

CogNet-KG: Empowering Tutoring Dialogues with a Cognitively-Structured Knowledge Graph for STEM Learning

Ding Yu¹, Yu Lu^{2,3}, Tengju Li¹, Shasha Xiong^{2,3}, Shengquan Yu^{2,3,*}

¹School of Artificial Intelligence, Beijing Normal University

²Advanced Innovation Center for Future Education, Beijing Normal University

³School of Educational Technology, Faculty of Education, Beijing Normal University

Correspondence: yusq@bnu.edu.cn

Abstract

Educational knowledge graph (EKG) is a critical component of intelligent tutoring systems that is structured around cognitive principles and provides support for interactive teaching. Most existing EKGs usually rely on simplistic relations, bind with single subjects, and lack integration with explicit learning objectives. In this paper, we introduce CogNet-KG, a novel and cognitively-structured large-scale knowledge graph for STEM learning. CogNet-KG models nearly 500 core concepts across five subjects with various cognitively-grounded relations corresponding to specific learning objectives, thereby encoding a rich cognitive schema for guiding more effective teaching. Based on this structure, we then construct a high-quality tutoring dialogue dataset CogDialogue-QA by leveraging adaptive instructional strategies. Additionally, we train CogTutor-LM, a specialized tutorial LLM that internalizes this structured pedagogical reasoning. Overall evaluation demonstrates that CogTutor-LM generates responses with significantly greater instructional coherence and more appropriate pedagogical guidance compared to baselines, validating the effectiveness of our graph-driven approach to fostering knowledge integration and stimulating students' thinking. The datasets are publicly available at <https://github.com/KCAIED/CogNet-KG>.

1 Introduction

In recent years, Large Language Models (LLMs) (Zhao et al., 2023) have been widely applied in various Intelligent Tutoring Systems (ITS) to enhance dialogue teaching due to its rich knowledge and impressive conversational fluency (Yan et al., 2024). However, a single LLM often lacks the structured and pedagogical reasoning required for guided learning such as diagnosing misconceptions and promoting conceptual transfer.

Therefore, Educational Knowledge Graphs (EKGs) (Qu et al., 2024) can reduce this limita-

tion by modeling curricular concepts, their relationships, and associated learning resources. A high-quality EKG serves as the foundational “cognitive map” for ITS, promoting critical downstream applications such as learning path recommendation (Wei et al., 2025), cognitive diagnosis (Gao et al., 2021) and knowledge tracing (Lu et al., 2022). The utility and intelligence of these applications are fundamentally correlated with the pedagogical fidelity and structural richness of the corresponding EKGs.

However, the development of such educational KGs remains a significant challenge. Current efforts usually focus on simple semantic connections between knowledge points that diminish their pedagogical value: 1) The relational schemas usually emphasize semantic (e.g., similarity) (Su and Zhang, 2020) or sequential links (e.g., prerequisite) (Chen et al., 2018) without capturing the nuanced cognitive relationships (e.g., expand, apply, compare) that underpin conceptual understanding and transfer, which results in lacking instructional logic and integrated knowledge structure; 2) Most EKGs are constructed for isolated subjects or grade levels and fail to model the essential cross-subject and cross-grade connections (Qu et al., 2024) that are crucial for building an integrated knowledge framework in subjects like integrated STEM; 3) Existing EKGs consist primarily of isolated triples and rarely integrate explicit learning objectives that indicate student' outcomes or behaviors in detail (Almatrafi and Johri, 2025), which leads to the absence of the link between abstract knowledge structures and concrete pedagogical actions and reduces their capability of guiding interactive teaching.

In this paper, to address these gaps, we introduce a novel framework that utilizes a cognitively-structured knowledge graph to generate diagnostic tutorial dialogue. First, we present a methodology for constructing CogNet-KG, which is an integrated EKG comprising nearly 500 core concepts across five STEM subjects (Mathematics,

Physics, Chemistry, Biology, Geography) spanning secondary school. We design a multi-agent framework with large language models (LLMs) playing several roles to extract and formalize not only standard relations but also pedagogically-grounded ones (e.g., transfer, apply, compare), and each relation corresponds to explicit learning objectives, which creates a rich and multi-relational educational knowledge graph. Second, we propose CogDialogue-QA, a high-quality dataset of simulated diagnostic tutoring dialogues by leveraging the structural and pedagogical information in CogNet-KG. The dialogues embody a pedagogical strategy where the teacher’s next question is determined by the student’s response and the cognitive relationships defined in the graph, which aims to simulate the behavior of high-level teachers such as remediation, consolidation, or extension. Third, we train CogTutor-LM on a base model using the fine-tuning method and CogDialogue-QA dataset, which serves as a dialogue agent whose feedback and questioning strategies are implicitly guided by CogNet-KG. Evaluations show that CogTutor-LM generates responses with significantly improved pedagogical coherence and more reasonable instructional decision compared to base models and general LLMs, which effectively promote the activation and construction of student knowledge structures.

The contributions of this paper are as follows:

- We propose CogNet-KG, a publicly available educational knowledge graph for STEM learning that is cognitively-structured, cross-grade/subject, and explicitly aligned with actionable learning objectives by incorporating a multi-agent framework.
- We construct the CogDialogue-QA dataset, where the tutoring dialogues between student and teacher are procedurally guided by the relational logic and learning objectives embedded in CogNet-KG, which aligns with the principles of tutoring.
- We present a fine-tuning model CogTutor-LM that internalizes the cognitive relationship and knowledge structure encoded in CogNet-KG, and it shows more educationally effective tutorial interactions according to the result of a comprehensive evaluation.

2 Related Work

2.1 Educational Knowledge Graphs

Educational knowledge graphs usually aim to construct internal logical relationships and map students’ thinking patterns and higher-order cognitive processes (Qu et al., 2024), which is an important foundation for organizing learning resources (You et al., 2024). With the development of technology, recent research tends to use automatic methods (e.g., text mining) to construct EKGs (Dang et al., 2021) from large and heterogeneous data (Zhao et al., 2022). For example, the BERT-BiLSTM-CRF model can be utilized to identify the subject knowledge points from educational data (Su and Zhang, 2020), while probabilistic association rule mining on learning assessment data can be applied to extract relationships (Chen et al., 2018). To further expand the modality of information, MEduKG (Li et al., 2022) fuses teacher speech data linking to the entities and then constructs a multi-modal knowledge graph. However, most constructions often rely on simple relational types (e.g., include, similar, prerequisite), which fails to reflect various cognitive processes (e.g., transfer (Haskell, 2001), problem-solving (Fuchs et al., 2004), learning objective (Almatrafi and Johri, 2025)) in real-world educational scene and make it difficult for ITS to utilize the content.

2.2 Large Language Models in Education

Large Language Models (LLMs) such as ChatGPT (Ouyang et al., 2022) have achieved great success due to their powerful natural language understanding and reasoning abilities (Zhao et al., 2023), and they can also handle complicated and diverse tasks in education (Yan et al., 2024). ReAL (Lv et al., 2025a) requires LLM to simulate teacher decision-making with extensive prior knowledge and teaching experience for adaptive learning. For question generation, EQPR (Cheng et al., 2025) formalizes it as a Markov Decision Process for deep reasoning and iterative improvement via the critic and reflection module, while CyberQ (Agrawal et al., 2024) designs a KG-augmented LLMs pipeline to generate question-answers by using different kinds of prompt-chaining. For teaching planning, EduPlanner (Zhang et al., 2025) is an LLM-based multi-agent system by working in adversarial collaboration to generate instructional design.

In interactive teaching, LLMs can produce appropriate responses based on student feedback by

integrating advanced teaching strategy such as scaffolding for problem-solving (Sonkar et al., 2023) and Socratic teaching (Peng et al., 2025) throughout the conversation. In addition, the paradigm of fine-tuning large language models on domain-specific data has shown great promise for enhancing the models’ ability to generate more specific and high-quality tutoring dialogues (Wang et al., 2025; Dan et al., 2023; Qi et al., 2025). However, this approach usually relies on unstructured or sequentially-ordered text corpora. To realize effective knowledge understanding and transfer, structured and relational knowledge information should be explicitly encoded as completely as possible into the dataset.

3 Method

3.1 Overview

CogNet-KG builds upon the SLP dataset (Lu et al., 2021), which contains a comprehensive system for K12 education domain knowledge labeled by experts. In the dataset, each knowledge point k_i has its own grade g_i , subject s_i , and theme t_i . It also has a set of learning objectives $\mathbf{L}_i = \{l_{i,1}, \dots, l_{i,|\mathbf{L}_i|}\}$, and each learning objective describes the behavior that students should achieve in detail, which is a high-quality corpus compared to other educational data. However, all of these knowledge contents are organized in a tree structure that fails to reflect the cognitive logical relations between knowledge points, which are more important and meaningful in real-world educational scenarios than isolated knowledge point. Thus, we define the task of extracting cognitive relations as follows: given the scope of all knowledge points \mathbf{K} , the set of all types of relation \mathbf{R} and a learning objective $l_{i,n}$ (with the side information k_i, g_i, s_i and t_i), our objective is to get a set of relevant relations $\mathbf{Rel}_{i,n} = \{\mathbf{m}_{i,n}^1, \dots, \mathbf{m}_{i,n}^{|\mathbf{Rel}_{i,n}|}\}$, where $\mathbf{m}_{i,n}^j = (k_{i,n}^{j,1}, r_{i,n}^j, k_{i,n}^{j,2})$ is a triple relation, $k_{i,n}^{j,1}, k_{i,n}^{j,2} \in \mathbf{K}$ and $r_{i,n}^j \in \mathbf{R}$. After processing all learning objectives, we can obtain a complete cognitively-structured knowledge graph called CogNet-KG. Our research focuses on STEM learning in secondary school because their knowledge systems are characterized by explicit cognitive and logical dependencies, which are more effective in formal representation in a structured knowledge graph compared to the humanities and social sciences.

Inspired by a large number of existing pedagogical and cognitive theories such as Bloom’s Taxonomy of Educational Objectives (Anderson et al., 2001), we summarize 9 types of relations that represent cognitive logic between knowledge points in collaboration with subject experts, which contain 7 directed relations and 2 undirected relations. Their definitions and examples can be found in Appendix A.

As shown in Figure 1, the construction of CogNet-KG follows a multi-stage analysis and design process to ensure the validity and effectiveness of the generated results, and the detailed prompt templates can be found in Appendix B. First, there is a preliminary assessment of the knowledge points in the learning objective. Based on the analysis, they align with the knowledge base \mathbf{K} to form candidate options. Finally, all reasonable cognitive relationships existing in the learning objective are inferred and are outputted in the form of a triplet list through a review process. For brevity, we omit the superscript i and n in the later method description.

3.2 Knowledge Recognizer (Agent_A)

Given the relatively large scale of the knowledge base, it is not appropriate to inform the LLM of all the knowledge points information. Thus, knowledge recognizer is responsible for initial understanding and recognition of the learning objective l , which aims to extract the key knowledge components in it:

$$\mathbf{a} = \text{Agent_A}(l, g, s, t, k) \quad (1)$$

Here, \mathbf{a} is a raw knowledge points list that contains several knowledge components.

3.3 Knowledge Selector (Agent_B)

The knowledge components in \mathbf{a} cannot be aligned with the knowledge base \mathbf{K} in terms of name, description, or granularity. Therefore, we design a retrieval module using text embedding to encode knowledge points both in \mathbf{K} and \mathbf{a} , then filter the top 10 most similar knowledge points (using cosine similarity) in \mathbf{K} to generate a candidate list \mathbf{c} for each knowledge component in \mathbf{a} . Later, knowledge selector performs a matching operation to determine the final knowledge point list \mathbf{L} :

$$\mathbf{L} = \text{Agent_B}(l, g, s, t, k, \mathbf{a}, \mathbf{c}) \quad (2)$$

Here, we can ensure that all knowledge points in \mathbf{L} do not exceed the scope of \mathbf{K} : $\mathbf{L} \subseteq \mathbf{K}$.

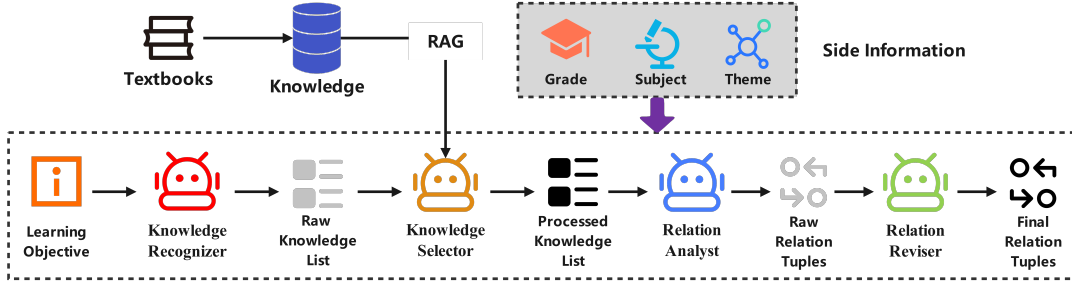


Figure 1: The multi-agent framework for constructing CogNet-KG.

3.4 Relation Analyst (Agent_C)

After confirming the knowledge point candidate list \mathbf{L} , the relation analyst aims to generate a raw set of relevant relations \mathbf{Rel}^r based on the given relation types \mathbf{R} :

$$\mathbf{Rel}_r = \text{Agent}_C(l, g, s, t, k, \mathbf{L}, \mathbf{R}) \quad (3)$$

Here, each $\mathbf{m}_r^j = (k_r^{j,1}, r_r^j, k_r^{j,2})$ in \mathbf{Rel}^r is a triple relation, where $k_r^{j,1}, k_r^{j,2} \in \mathbf{L}$ and $r_r^j \in \mathbf{R}$.

3.5 Relation Reviser (Agent_D)

Due to the hallucination of LLMs, we add an additional review mechanism to improve the quality of the final results. The relation reviser makes a judgment on each relation \mathbf{m}_r^j to determine whether it should be deleted, retained, or reversed. Therefore, we can obtain the final relations:

$$\text{Agent}_D : \mathbf{Rel}_r \rightarrow \mathbf{Rel} \quad (4)$$

It is worth mentioning that sometimes there may be a situation $|\mathbf{L}| \leq 1$, which indicates that the learning objective l is only within the knowledge point k without connecting to other knowledge points. Thus, as shown in Figure 2, we set a special relation type "self-correlation" to represent the information, which can maintain uniformity that all learning objectives are marked on the edges of CogNet-KG. More examples in CogNet-KG can be found in Appendix C.

4 The Datasets and the Fine-Tuning Model

4.1 CogNet-KG

After automated construction, we invite experienced teachers in corresponding subjects to review CogNet-KG by determining whether each relation should be retained, reversed or deleted based on the definition. The final results of the three options

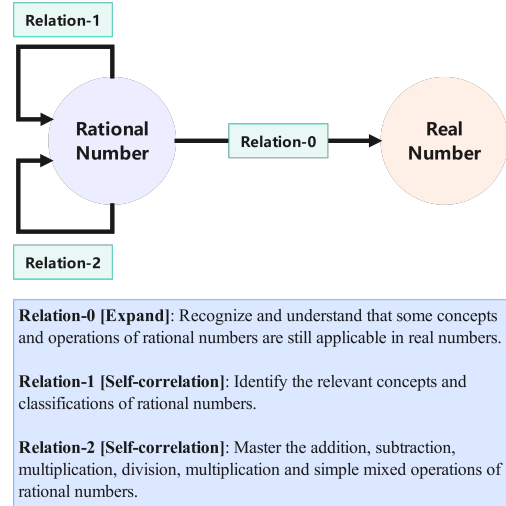


Figure 2: A subgraph of CogNet-KG, where each relation is marked with its type and learning objective.

is 87%, 6%, and 7%, which demonstrates the effectiveness of the multi-agent framework and confirms the final graph. Later, we conduct a comprehensive statistical analysis of the graph to show its potential to support tutoring, and more results can be found in Appendix D.

4.1.1 Quantitative Study

In CogNet-KG, there are 494 knowledge points as nodes, 7920 relations as edges (2574 relations excluding self-correlation type). Table 3 shows the proportion of various relation types in each combination of grades and subjects, which intuitively reflects some typical features of them. For example, all subjects emphasize the ability to apply knowledge to real-world problems; the required cognitive logic gradually diversifies with the increasing of grade; The reasoning ability required for mathematics, physics and chemistry is relatively high; biology and geography subjects tend to examine students through comparative tasks based on case studies.

4.1.2 Graph Analysis

Besides the basic data statistics, we utilize some graph algorithm to explore the structural features of CogNet-KG, which provide effective guidance for educational scenario.

- **Community Detection:** We utilize the Louvain algorithm to detect critical communities in each combination of grades and subjects. As the examples shown in Table 10 and Table 11, similar knowledge points tend to be classified into the same community, which can serve as the reference for **large unit teaching** or **big idea teaching**.
- **Betweenness Centrality:** We use this indicator to calculate the key propagation channels in CogNet-KG. As the examples shown in Table 12, they can be regarded as **the keynotes of review and consolidation** in practical teaching.
- **Cross-grade and Cross-subject Analysis:** Table 13 and Table 14 provide some typical knowledge points in these kinds of relations, which can provide effective guidance for teachers to launch **the transition from junior to senior secondary school** and **project-based learning**.

4.2 CogDialogue-QA

Based on the high-quality educational knowledge CogNet-KG, we further construct a dialogue dataset (CogDialogue-QA) by simulating conversations between teacher and students, which aims to transform the structured and pedagogical logic of CogNet-KG into natural language interactions.

In particular, we use GPT-4o to implement a rule-based dialogue generation process where the teacher’s next instruction is determined by mapping the student’s proficiency level (mastered, partially mastered, unmastered) to specific relational traversals within CogNet-KG, which directly reflects the educational principle of Zone of Proximal Development (Vygotsky, 1978) and adaptive scaffolding (Hammond, 2001).

Specifically, each relation $m = (k_1, r, k_2)$ (with its learning objective l) in CogNet-KG \mathcal{G} is assigned four relation sets:

Precursor (PRE): The set represents all the precursor relations of m , which indicates the fundamental preparation for the learning objective. As

shown in Equation 5, there are different rules for different relation types.

Posterior (POST): The set represents all the successor relations of m , which indicates the advanced development for the learning objective. As shown in Equation 6, there are different rules for different relation types.

Similarity (SIM): The set represents all the similar relations of m , which indicates slight adjustments for the learning objective. As shown in Equation 7, this set is only applicable to relations belonging to self-correlation type.

Self-Correlation (SELF): The set represents the own knowledge points of m , which indicates review for the present learning objective. As shown in Equation 8, this set is only applicable to relations belonging to self-correlation type.

Furthermore, we set different types of simulated students based on their mastery level and then determine different tutoring strategies. For a student demonstrating good **mastery**, the teacher tends to lateral or forward relations to promote knowledge transfer and deepening. For **partial mastery**, the teacher’s focus shifts to reinforcing core concepts and their immediate associations to solidify coherence. For **non-mastery**, the teacher should navigate prerequisite relationships to diagnose and address foundational gaps. Therefore, the candidate set of next learning objectives for different students and relation types is shown in Table 1. The prompt template for constructing CogDialogue-QA can be found in Appendix E, while the example of the dialogue dataset can be found in Appendix F.

The above process ensures that each dialogue turn is not merely a fluent text continuation, but a pedagogically intentional step guided by the cognitive structure to enhance conceptual understanding and transfer. The CogDialogue-QA consists of 23209 teacher-student interaction turns, covering all relations with their learning objectives in CogNet-KG. This dialogue dataset provides a solid training foundation for CogTutor-LM to internalize the instructional reasoning and support adaptive learning (Lv et al., 2025b).

4.3 CogTutor-LM

Since CogDialogue-QA is a Chinese dataset, we train CogTutor-LM on Qwen-3-8B using the LoRA fine-tuning method, with 3 epochs, a learning rate of $5e^{-5}$, a batch size of 16, a LoRA rank of 16, a LoRA alpha of 32, and a LoRA dropout of 0.1. The dataset is split into 90% for training and 10%

$$PRE = \begin{cases} \{m_r \in \mathcal{G} \mid m_r.directionality = \text{directed} \ \& \ m_r.k_2 = k_1 \ \& \ m_r \neq m\}, & r = \text{self_correlation}; \\ \{m_r \in \mathcal{G} \mid m_r.r = \text{self_correlation} \ \& \ m_r.k_2 = k_1 \ \& \ m_r \neq m\}. & \textit{else.} \end{cases} \quad (5)$$

$$POST = \begin{cases} \{m_r \in \mathcal{G} \mid m_r.directionality = \text{directed} \ \& \ m_r.k_1 = k_2 \ \& \ m_r \neq m\}, & r = \text{self_correlation}; \\ \{m_r \in \mathcal{G} \mid m_r.r = \text{self_correlation} \ \& \ m_r.k_1 = k_2 \ \& \ m_r \neq m\}. & \textit{else.} \end{cases} \quad (6)$$

$$SIM = \{m_r \in \mathcal{G} \mid m_r.directionality = \text{undirected} \ \& \ (m_r.k_1 = k_2 \ || \ m_r.k_2 = k_1) \ \& \ m_r \neq m\} \quad (7)$$

$$SELF = \{m_r \in \mathcal{G} \mid m_r.r = \text{self_correlation} \ \& \ m_r.k_1 = k_1 \ \& \ m_r.k_2 = k_2 \ \& \ m_r \neq m\} \quad (8)$$

	Mastery	Partial Mastery	Non-Mastery
Self-correlation	SIM, POST	SIM, SELF	PRE, SELF
Directed	POST	PRE, POST	PRE
Undirected	PRE, POST	PRE, POST	PRE, POST

Table 1: The candidate list of next learning objectives in different cases.

for testing. It should be noted that we only set the content of teacher-ask ta and student-answer sa as the input and set the content of teacher-tutoring tt as the expected output. In other words, we hide the explicit information of CogNet-KG from the LLMs that maximizes the following conditional probability:

$$P_{LLMs}(tt|ta \oplus sa; \Phi) \quad (9)$$

where Φ are the parameters of the LLMs and \oplus represents the string concatenation operation. Through fine-tuning based on the CogDialogue-QA dataset, CogTutor-LM acquires the capability to conduct tutoring dialogues including follow-up questions and explanations that are implicitly guided by the cognitive structure of CogNet-KG, which helps improve instructional coherence and foster targeted knowledge activation.

5 Evaluation

5.1 Tutorial Interaction Quality Evaluation System

We evaluate our CogTutor-LM using both automatic metrics and human assessment. For automatic evaluation, we first select widely-adopted natural language generation (NLG) metrics to comprehensively assess text quality, such as Rouge-1 (R-1), Rouge-2 (R-2), Rouge-L (R-L), and BLEU-4. However, these standard metrics are inadequate for our task, as they primarily assess surface-level lexical overlap, which fails to capture the pedagogical quality and cognitive rationality of a tutoring

response. For example, a response could be fluent and lexically rich but pedagogically misguided. Therefore, we propose a tutorial interaction quality evaluation system for LLMs, which covers nine dimensions raised by domain experts and each dimension is assessed on a 5-point scale for GPT-4o to evaluate (more details can be found in Appendix G). This framework provides a valid and reliable assessment of whether the LLMs truly leverage the cognitively-structured knowledge graph to make educationally sound decisions.

Our tutorial interaction quality evaluation system can be divided into three levels: instructional content, interactive support, and individual development. The details are as follows.

Instructional Content: These dimensions focus on whether the organization, expression, and structural design of teaching content are clear, coherent, and accord with cognitive laws.

- **Clarity (A1):** Evaluates the legibility, specificity, and appropriateness of teaching objectives and content expression.
- **Expression (A2):** Evaluates the organization logic and matching degree of teaching language with students’ cognition and knowledge proficiency.
- **Coherence (A3):** Evaluates the naturalness and logical flow of contextual memory and connection within the dialogue.
- **Constructiveness (A4):** Evaluates the degree to which the logical progression and cognitive

	R-1	R-2	R-L	BLEU-4	A1	A2	A3	A4	B1	B2	B3	C1	C2
GPT-4o	27.92	6.82	21.77	4.16	<u>3.73</u>	<u>3.73</u>	3.38	3.49	3.58	<u>3.49</u>	3.39	2.51	3.08
GPT-4o-RAG	24.36	5.99	20.87	3.27	3.72	3.69	3.25	3.36	3.59	3.37	3.32	2.49	3.10
Llama-3-8B	23.47	4.12	18.42	2.71	3.58	3.69	3.23	3.05	3.31	2.84	3.00	2.21	2.87
GLM-4-9B	28.78	6.51	21.89	4.00	3.71	3.71	3.46	3.49	3.56	3.34	<u>3.40</u>	2.55	3.20
GLM-3-6B	28.28	6.80	<u>21.96</u>	4.12	3.69	3.69	3.34	3.40	3.59	3.43	3.36	<u>2.57</u>	3.05
Qwen-2.5-7B	26.47	5.51	20.13	3.17	3.71	3.71	3.36	3.41	3.43	3.20	3.26	2.37	2.97
Qwen-3-8B	26.67	5.98	20.40	3.78	3.71	3.70	3.29	3.42	<u>3.60</u>	3.27	3.31	2.34	3.11
Educhat-13B	<u>30.17</u>	<u>8.47</u>	21.16	3.95	3.68	3.70	<u>3.47</u>	3.25	3.24	3.05	3.21	2.24	3.28
CogTutor-LM	41.93	15.34	32.69	9.45	3.80	3.85	3.67	3.60	3.62	3.59	3.44	2.61	<u>3.27</u>

Table 2: The performances of tutorial interaction quality. For all metrics, higher values indicate better performance. The top-performing are bolded, and the second-best are underlined.

connection between knowledge points across the dialogue facilitate systematic, structured knowledge building.

Interactive Support: These dimensions focus on the effectiveness of interactive strategies used by teachers in identifying student states, guiding thinking, and providing feedback during tutoring.

- **Diagnosis (B1):** Evaluates the accuracy in identifying students’ knowledge state, core misunderstandings, and cognitive obstacles.
- **Guidance (B2):** Evaluates the ability to stimulate thinking, promote reflection, and support autonomous learning planning through questioning or prompting.
- **Feedback (B3):** Evaluates the accuracy, constructiveness, and specificity of explanations, evaluations, and improvement suggestions in response to errors.

Individual Development: These dimensions focus on whether the teaching takes into account individual differences and knowledge states to activate the learning motivation of students and support their personalized growth.

- **Personalization (C1):** Evaluates the ability to tailor learning paths and content based on student readiness, cognitive state or style.
- **Motivation (C2):** Evaluates the effectiveness in boosting student confidence and intrinsic motivation through emotional interaction and attribution style.

5.2 Evaluation of CogTutor-LM

We select several strong-performing models (with similar parameter scales) on Chinese tasks as baselines to evaluate the effectiveness. These include GPT-4o, GPT-4o-RAG (retrieving the top 5 most

relevant learning objectives based on the present conversation), Llama-3-8B (Touvron et al., 2023), GLM-4-9B, GLM-3-6B (GLM et al., 2024), Qwen-2.5-7B (Qwen et al., 2025), Qwen-3-8B (Yang et al., 2025), and Educhat-13B (Dan et al., 2023) (known for incorporating theories of psychology and education). To avoid the model from traversing all relevant content and catering to metrics coincidentally, we uniformly limit the max tokens of the output scale to 128, which also meets the requirements of concise and refined language in educational dialogue scenarios. To verify the reliability of the automated evaluations by GPT-4o, we randomly select 100 samples from the test set and invite 4 experts with relevant educational background to evaluate them. The ICC coefficients are as follows: Clarity (0.69), Expression (0.72), Coherence (0.81), Constructiveness (0.78), Diagnosis (0.65), Guidance (0.79), Feedback (0.67), Personalization (0.61) and Motivation (0.74), indicating a high level of consistency between GPT-4o and human evaluations. Then, Table 2 summarizes the performance of all models across the evaluation metrics. In general, our CogTutor-LM demonstrates superior performance in tutorial interaction.

For standard NLG metrics, CogTutor-LM outperforms due to our fine-tuning work, while Educhat-13B is generally second best, which indicates that the dialogues in CogDialogue-QA basically follow the teaching principles implied in Educhat.

For the metrics in tutorial interaction quality evaluation system, CogTutor-LM has significant advantages especially in Expression, Coherence, Constructiveness and Guidance, which indicates that it has a good grasp of the cognitive structure information in CogNet-KG and can effectively promote students’ conceptual understanding and transfer. However, the improvement of CogTutor-LM is relatively small in Diagnosis and Motivation, for we have not yet focused on the relevant features dur-

<p>[Teacher-ask]: Today we will review how to use number lines and absolute values to compare the size of rational numbers. The numbers on the number line increase as they move to the right and decrease as they move to the left, while the absolute value represents the distance from a number to the origin, which can help us determine its size. Please think and answer: Which number is greater on the number line, -3 or 2? How to compare the absolute values of them? [Student-answer]: -3 is closer to the left than 2 on the number line, so it is smaller than 2; Their absolute values are 3 and 2, obviously 3 is greater than 2.</p>		
<p>GPT-4o: The students' answers are generally correct, able to accurately determine the size relationship between -3 and 2 on the number line, and correctly compare their absolute values. But further guidance can be given to students to clearly express the meaning of absolute values, such as adding that "absolute value represents the distance from the number to the origin, so the absolute value of -3 is 3, the absolute value of 2 is 2, and 3 is greater than 2". This can help students understand concepts more clearly and improve their communication skills.</p>	<p>Educhat-13B: Your answer is basically correct, but you need to explain the concept of absolute value in more detail. The absolute value represents the distance from a number to the origin, so the absolute values of -3 and 2 are 3 and 2, respectively, and 3 is larger than 2, so -3 is smaller than 2. Additionally, you can use a number line to visually compare the size of two numbers, placing them on the line to see which number is closer to the right. This can provide a more intuitive understanding of the concepts of number axis and absolute value. I hope you can better grasp these knowledge points.</p>	<p>CogTutor-LM: Very good, you have fully mastered how to compare the size of rational numbers using number lines and absolute values. Next, we will explore the relationship between rational and irrational numbers, and how they together form real numbers. Please consider: What are the differences in the rules of operation between rational and irrational numbers?</p>

Table 3: The tutoring responses generated by different LLMs.

ing the construction of the dataset, and most base LLMs are sufficient to meet such requirements.

GPT-4o is usually second best, which has a much larger parameter size and exhibits remarkable fundamental capabilities such as wide extensive knowledge and comprehension of language. In addition, RAG is driven by semantic similarity and cannot select the most suitable learning objective based on the cognitive structure, which results in misleading effects.

5.3 Case Study

To further demonstrate the actual effectiveness of this model, we randomly select a sample from the testing set to conduct a case study. We choose GPT-4o and Educhat-13B for comparison, which perform relatively well in the above experiment.

As shown in Table 3: GPT-4o tends to provide a more detailed explanation of the learning objective itself although the student has answered the question completely correct; The response generated by Educhat-13B is more vivid and informative, but the learning suggestions provided still lack inspiration and development. For CogTutor-LM, although

its teaching dialogue is shorter in length, it generally covers the evaluation and encouragement to student's answers, clear further learning objective based on the knowledge structure and effective questioning to stimulate student's thinking, which has stronger practical educational significance.

6 Conclusion

In this paper, to enhance knowledge representation and effective tutoring dialogues in STEM education, we propose a cognitively-structured knowledge graph CogNet-KG by designing a multi-agent framework. CogNet-KG integrates various cognitive relations and spans multiple grades and subjects, each relation explicitly corresponding to an actionable learning objective. Besides, we use the rich information in CogNet-KG and design structured pedagogical reasoning to generate a high-quality tutorial dialogue dataset, CogDialogue-QA, based on which we train CogTutor-LM serving as a teacher agent that internalizes instructional logic. Experimental results show that CogTutor-LM generates more coherent and pedagogically sound dialogue turns than general LLMs.

Limitations

Our work has still several limitations. First, CogNet-KG is currently confined to STEM subjects, while the generalizability of our approach to humanities subjects remains unverified. Second, CogTutor-LM generally outperforms baselines, but its performance on some metrics like diagnosis or motivation can be further improved. Furthermore, although CogNet-KG itself is cross-grade and cross-subject, our dialogue dataset generation and model fine-tuning processes do not explicitly design specific mechanisms to leverage or evaluate these important features. Future research should aim to expand the knowledge paradigm to include diverse subjects, refine the model’s diagnostic reasoning through more advanced data processing and training techniques to adapt to multi-turn tasks such as knowledge tracing and Socratic teaching, and also explicitly explore how cross-grade or cross-subject relation can be exploited to promote more integrated and transferable learning.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (No. 62477002). We thank the anonymous reviewers for the valuable comments.

References

- Garima Agrawal, Kuntal Pal, Yuli Deng, Huan Liu, and Ying-Chih Chen. 2024. [CyberQ: Generating Questions and Answers for Cybersecurity Education Using Knowledge Graph-Augmented LLMs](#). *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(21):23164–23172.
- Omaima Almatrafi and Aditya Johri. 2025. [Leveraging generative ai for course learning outcome categorization using bloom’s taxonomy](#). *Computers and Education: Artificial Intelligence*, 8:100404.
- Lorin W Anderson, David R Krathwohl, Peter W Airasian, Kathleen A Cruikshank, Richard E Mayer, Paul R Pintrich, James Raths, and Merlin C Witrock. 2001. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom’s Taxonomy of Educational Objectives*. Longman, New York, NY.
- Penghe Chen, Yu Lu, Vincent W. Zheng, Xiyang Chen, and Boda Yang. 2018. [KnowEdu: A System to Construct Knowledge Graph for Education](#). *IEEE Access*, 6:31553–31563.
- Cheng Cheng, Zhenya Huang, GuanHao Zhao, Yuxiang Guo, Xin Lin, Jinze Wu, Xin Li, and Shijin Wang. 2025. [From objectives to questions: A planning-based framework for educational mathematical question generation](#). In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12836–12856, Vienna, Austria. Association for Computational Linguistics.
- Yuhao Dan, Zhikai Lei, Yiyang Gu, Yong Li, Jianghao Yin, Jiaju Lin, Linhao Ye, Zhiyan Tie, Yougen Zhou, Yilei Wang, Aimin Zhou, Ze Zhou, Qin Chen, Jie Zhou, Liang He, and Xipeng Qiu. 2023. [EduChat: A Large-Scale Language Model-based Chatbot System for Intelligent Education](#).
- Fu-Rong Dang, Jin-Tao Tang, Kun-Yuan Pang, Ting Wang, Sha-Sha Li, and Xiao Li. 2021. [Constructing an Educational Knowledge Graph with Concepts Linked to Wikipedia](#). *Journal of Computer Science and Technology*, 36(5):1200–1211.
- Lynn S. Fuchs, Douglas Fuchs, Karin Prentice, Carol L. Hamlett, Robin Finelli, and Susan J. Courey. 2004. [Enhancing mathematical problem solving among third-grade students with schema-based instruction](#). *Journal of Educational Psychology*, 96(4):635–647.
- Weibo Gao, Qi Liu, Zhenya Huang, Yu Yin, Haoyang Bi, Mu-Chun Wang, Jianhui Ma, Shijin Wang, and Yu Su. 2021. [RCD: Relation map driven cognitive diagnosis for intelligent education systems](#). In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Sigir ’21*, pages 501–510, New York, NY, USA. Association for Computing Machinery.
- Team GLM, :, Aohan Zeng, Bin Xu, Bowen Wang, Chenhui Zhang, Da Yin, Dan Zhang, Diego Rojas, Guanyu Feng, Hanlin Zhao, Hanyu Lai, Hao Yu, Hongning Wang, Jiadao Sun, Jiajie Zhang, Jiale Cheng, Jiayi Gui, Jie Tang, and 40 others. 2024. [Chatglm: A family of large language models from glm-130b to glm-4 all tools](#). *Preprint*, arXiv:2406.12793.
- Ed Hammond, Jennifer. 2001. *Scaffolding: Teaching and Learning in Language and Literacy Education*. Scaffolding: Teaching and Learning in Language and Literacy Education.
- R. E. Haskell. 2001. *Transfer of learning: Cognition, instruction, and reasoning*. Academic Press.
- Nan Li, Qiang Shen, Rui Song, Yang Chi, and Hao Xu. 2022. [MEduKG: A deep-learning-based approach for multi-modal educational knowledge graph construction](#). *Information-an International Interdisciplinary Journal*, 13(91).
- Yu Lu, Penghe Chen, Yang Pian, and Vincent W. Zheng. 2022. [CMKT: Concept Map Driven Knowledge Tracing](#). *IEEE Transactions on Learning Technologies*, 15(4):467–480.
- Yu Lu, Yang Pian, Ziding Shen, Penghe Chen, and Xiaoqing Li. 2021. [SLP: A Multi-Dimensional and](#)

- Consecutive Dataset from K-12 Education. In *29TH INTERNATIONAL CONFERENCE ON COMPUTERS IN EDUCATION (ICCE 2021), VOL 1*, pages 261–266, NO 300, JUNGDA RD, JHONGLI DISTRICT, TAOYUAN CITY, 320, TAIWAN. ASIA PACIFIC SOC COMPUTERS IN EDUCATION.
- Rui Lv, Qi Liu, Weibo Gao, Jiatong Li, Kai Zhang, and Shiwei Tong. 2025a. [ReAL: How can LLMs simulate the real teacher? Retrieval-enhanced agent for adaptive learning](#). In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 168–181, Suzhou, China. Association for Computational Linguistics.
- Rui Lv, Qi Liu, Weibo Gao, Jiatong Li, Kai Zhang, and Shiwei Tong. 2025b. [ReAL: How can LLMs simulate the real teacher? Retrieval-enhanced agent for adaptive learning](#). In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 168–181, Suzhou, China. Association for Computational Linguistics.
- Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022. [Training language models to follow instructions with human feedback](#). *Preprint*, arXiv:2203.02155.
- Xian Peng, Pan Yuan, Dong Li, Junlong Cheng, Qin Fang, and Zhi Liu. 2025. [KELE: A multi-agent framework for structured socratic teaching with large language models](#). In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 16342–16362, Suzhou, China. Association for Computational Linguistics.
- Changyong Qi, Yuang Wei, Haoxin Xu, Longwei Zheng, Peiji Chen, and Xiaoqing Gu. 2025. [TMATH a dataset for evaluating large language models in generating educational hints for math word problems](#). In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 5082–5093, Abu Dhabi, UAE. Association for Computational Linguistics.
- Kechen Qu, Kam Cheong Li, Billy T. M. Wong, Manfred M. F. Wu, and Mengjin Liu. 2024. [A survey of knowledge graph approaches and applications in education](#). *Electronics*, 13(2537).
- Qwen, :, An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiayi Yang, Jingren Zhou, and 25 others. 2025. [Qwen2.5 technical report](#). *Preprint*, arXiv:2412.15115.
- Shashank Sonkar, Naiming Liu, Debshila Mallick, and Richard Baraniuk. 2023. [CLASS: A design framework for building intelligent tutoring systems based on learning science principles](#). In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 1941–1961, Singapore. Association for Computational Linguistics.
- Ying Su and Yong Zhang. 2020. [Automatic Construction of Subject Knowledge Graph based on Educational Big Data](#). In *Proceedings of the 2020 3rd International Conference on Big Data and Education, ICBDE '20*, pages 30–36, New York, NY, USA. Association for Computing Machinery.
- Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. [Llama: Open and efficient foundation language models](#). *Preprint*, arXiv:2302.13971.
- L. S. Vygotsky. 1978. *Mind in Society: Development of Higher Psychological Processes*. Harvard University Press.
- Haochun Wang, Sendong Zhao, Zewen Qiang, Zijian Li, Chi Liu, Nuwa Xi, Yanrui Du, Bing Qin, and Ting Liu. 2025. [Knowledge-tuning large language models with structured medical knowledge bases for trustworthy response generation in chinese](#). *ACM Trans. Knowl. Discov. Data*, 19(2).
- Xin Wei, Wenrui Han, Shiyun Sun, Junhao Shan, and Liang Zhou. 2025. [Educational psychology-empowered personalized learning path generation strategy](#). *IEEE Transactions on Learning Technologies*, 18:741–756.
- Lixiang Yan, Lele Sha, Linxuan Zhao, Yuheng Li, Roberto Martinez-Maldonado, Guanliang Chen, Xinyu Li, Yueqiao Jin, and Dragan Gašević. 2024. [Practical and ethical challenges of large language models in education: A systematic scoping review](#). *British Journal of Educational Technology*, 55(1):90–112.
- An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, and 41 others. 2025. [Qwen3 technical report](#). *Preprint*, arXiv:2505.09388.
- Weihao You, Pengcheng Wang, Changlong Li, Zhi-long Ji, and Jinfeng Bai. 2024. [CK12: A Rounded K12 Knowledge Graph Based Benchmark for Chinese Holistic Cognition Evaluation](#). *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(17):19431–19439.
- Xueqiao Zhang, Chao Zhang, Jianwen Sun, Jun Xiao, Yi Yang, and Yawei Luo. 2025. [EduPlanner: LLM-based multiagent systems for customized and intelligent instructional design](#). *IEEE Transactions on Learning Technologies*, 18:416–427.

Bowen Zhao, Jiuding Sun, Bin Xu, Xingyu Lu, Yuchen Li, Jifan Yu, Minghui Liu, Tingjian Zhang, Qiuyang Chen, Hanming Li, Lei Hou, and Juanzi Li. 2022. *EDUKG: A heterogeneous sustainable K-12 educational knowledge graph*. *Preprint*, arXiv:2210.12228.

Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Zhipeng Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, and 3 others. 2023. *A Survey of Large Language Models*.

A Explanation of the Relations

Table 4 shows the name, directionality, definition and example of all relation types we set.

B Prompt Templates for Knowledge Graph Construction

Table 5, Table 6, Table 7, and Table 8 show the prompt templates of knowledge recognizer, knowledge selector, relation analyst, and relation reviser, respectively.

C More Examples of CogNet-KG

Table 9 shows more relations and their corresponding learning objectives in CogNet-KG.

D More Analysis of CogNet-KG

In this section, we display some key analysis results of CogNet-KG (Figure 3, Table 10, Table 11, Table 12, Table 13, Table 14), and more features of the graph remain to be explored and excavated.

E Prompt Templates for Constructing CogDialogue-QA Dataset

Table 15 shows the prompt template for CogDialogue-QA construction.

F Example for the CogDialogue-QA Dataset

Table 16 shows an entire record in CogDialogue-QA including the learning objective, candidate list and teacher-student dialogue.

G The Scale of the Tutorial Interaction Quality Evaluation System

Table 17, Table 18, and Table 19 show the scales for our tutorial interaction quality evaluation system in three levels.

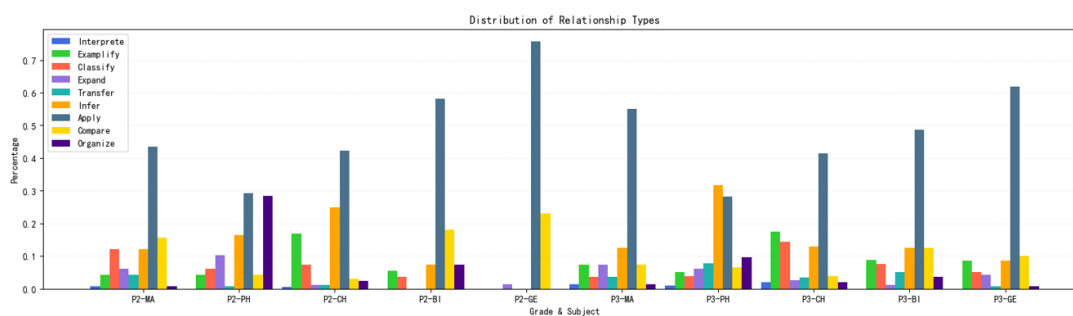


Figure 3: The proportion of various relation types in each grade-subject, where P2 and P3 represent the junior secondary school and senior secondary school stage respectively, and MA, PH, CH, BI and GE represent Mathematics, Physics, Chemistry, Biology and Geography respectively.

Relation	Directionality	Definition	Example
Interprete	Directed	The latter knowledge point is another advanced form of representation of the former knowledge point.	Description of Motion - Displacement [P3-PH] → Vectors [P3-MA]
Exemplify	Directed	The latter knowledge point is a relatively abstract concept, while the latter knowledge point is a more specific expression, context, case, experimental activity, etc.	Acids and Bases [P2-CH] → Experiment: Chemical Properties of Acids and Bases [P2-CH]
Classify	Directed	The latter knowledge point is a component or subclass of the former knowledge point.	Hydrocarbon [P3-CH] → Alkane [P3-CH]
Expand	Directed	The latter knowledge point can be obtained by expanding the scope or condition of the former knowledge point.	Rational Number [P2-MA] → Real Number [P2-MA]
Transfer	Directed	The latter knowledge point is obtained by adaptively adjusting the former knowledge point based on corresponding contextual requirements.	Congruent Triangle [P2-MA] → Similar Triangle [P2-MA]
Infer	Directed	The latter knowledge point can be connected to the former knowledge point through a logical chain.	Buoyancy and Archimedes principle [P2-PH] → The Floating and Sinking Conditions of Objects and Their Application [P2-PH]
Apply	Directed	The former knowledge point refers to relevant tools, clues, or conditions, while the latter knowledge point refers to a certain task, program, or knowledge.	Map [P2-GE] → Recognizing Continent [P2-GE]
Compare	Undirected	There are rich similarities and differences between two knowledge points that need to be compared and analyzed.	Photosynthesis (preliminary) [P2-BI] ↔ Respiratory function (preliminary) [P2-BI]
Organize	Undirected	The two knowledge points need to be integrated to establish an advanced relationship and model.	Law of Universal Gravitation [P3-PH] ↔ Circular Motion [P3-PH]

Table 4: The definition and examples of all relation types. The grade and subject of each knowledge point are marked after its name, where P2 and P3 represent the junior secondary school and senior secondary school respectively, and MA, PH, CH, BI and GE represent Mathematics, Physics, Chemistry, Biology and Geography respectively.

Knowledge Recognizer Prompt
<p>You are a secondary school subject expert, and you need to extract all the knowledge points as comprehensively and completely as possible based on the input learning objective (with grade, subject, theme, and concept for reference), and return them in a list format.</p> <p>Task Requirements:</p> <ol style="list-style-type: none"> 1. Ensure that the extracted knowledge points are complete, independent, and readable. Knowledge points also contain specific procedures or methods if necessary, such as "determining congruent triangles". 2. Please carefully analyze and understand the essential meaning of the learning objective, and avoid being misled by the local vocabulary in the original text. 3. Return a list as the result without any other content. <p>Example: {few_shots}</p> <p>Based on the task description and examples as mentioned, please extract the list of knowledge points included in the following learning objective.</p> <p>Grade and Subject: {grade} {subject} Theme: {theme} Concept: {concept} Learning Objective: {learning objective}</p>

Table 5: Knowledge recognizer prompt template.

Knowledge Selector Prompt
<p>You are a secondary school subject expert, and you need to understand the given knowledge points and learning objective (with grade, subject, theme, and concept for reference), and select the index number of the most relevant one for each knowledge point from the candidate list.</p> <p>Task Requirements:</p> <ol style="list-style-type: none"> 1. If you think there are no suitable option (e.g., mismatched academic subject, low relevance to the learning objective), please return "-1". 2. No adjustments or modifications shall be made to the options in the candidate list. 3. Return a list (with one index number for each knowledge point) as the result without any other content. <p>Based on the task description as mentioned, please match the options for the knowledge points.</p> <p>Grade and Subject: {grade} {subject} Theme: {theme} Concept: {concept} Learning Objective: {learning objective} Knowledge Points List: {knowledge points list} Candidate List: {candidate list}</p>

Table 6: Knowledge selector prompt template.

Relation Analyst Prompt
<p>You are a secondary school subject expert, and you need to understand the candidate knowledge points and learning objective (with grade, subject, theme, and concept for reference), and extract all the cognitive logical relations between them. The relations must come from the content of the learning objective, avoiding adding any extra relation type, and the selected knowledge points must come from the candidate list. Please note that the content of learning objective may sometimes be broad, but it is deeply bound to the reference information. Please fully understand all the content based on this side information, and avoid labeling irrelevant relations.</p> <p>Relation Types: {all_relations}</p> <p>Task Requirements:</p> <ol style="list-style-type: none"> 1. Please carefully analyze and understand the meaning of the learning objective, and avoid being misled by the local vocabulary in the original text. 2. The results are returned in the form of a two-dimensional list, and it is ensured to use the corresponding index numbers of the knowledge points in the candidate list. Do not output any other content except for this. 3. Each relation must use one of the above types, and no new relation type can be added, nor can their description be modified. <p>Based on the task description as mentioned, please analyze the following learning objective.</p> <p>Grade and Subject: {grade} {subject}</p> <p>Theme: {theme}</p> <p>Concept: {concept}</p> <p>Learning Objective: {learning objective}</p> <p>Knowledge Point Candidate List: {knowledge point candidate list}</p>

Table 7: Relation analyst prompt template. The content of "all_relations" is completely cited from Table 4.

Relation Reviser Prompt
<p>You are a secondary school subject expert, and you need to determine whether the extracted relations are consistent with the given learning objective (with grade, subject, theme, and concept for reference). You can make your decision based on the subject knowledge, common sense and the given relation types.</p> <p>The result of each relation can be divided into three situations:</p> <p>0: The relation is unreasonable and needs to be deleted.</p> <p>1: The relation is reasonable and should be retained.</p> <p>2: The relation is reasonable, but the order of the knowledge points should be reversed to meet the definition of the corresponding relationship type.</p> <p>Relation Types: {all_relations}</p> <p>Task Requirements:</p> <p>You only need to return one number from 0, 1, or 2 for each relation, and do not output anything else. Do not output any other content except for this.</p> <p>Based on the task description as mentioned, please analyze whether the extracted relations are reasonable.</p> <p>Knowledge Point A: {grade-A} {subject-A}{theme-A}{concept-A}{knowledge point-A}</p> <p>Relation: {relation}</p> <p>Knowledge Point B: {grade-B} {subject-B}{theme-B}{concept-B}{knowledge point-B}</p> <p>Learning Objective: {learning objective}</p>

Table 8: Relation reviser prompt template. The content of "all_relations" is completely cited from Table 4.

Relation Type and its Learning Objective
Interprete: Description of Motion - Displacement [P3-PH] → Vectors [P3-MA] Realize the concept of vectors and distinguish whether a physical quantity is a vector or a scalar.
Exemplify: Acids and Bases [P2-CH] → Experiment: Chemical Properties of Acids and Bases [P2-CH] Design experiments to demonstrate the typical properties of unfamiliar acids and bases.
Classify: Hydrocarbon [P3-CH] → Alkane [P3-CH] Compare the structures of ethylene and ethane and demonstrate their property differences based on functional groups.
Expand: Rational Number [P2-MA] → Real Number [P2-MA] Recognize and understand that some concepts and operations of rational numbers are still applicable in real numbers.
Transfer: Congruent Triangle [P2-MA] → Similar Triangle [P2-MA] Use the properties or judgment theorems of congruent triangles to infer the properties of similar triangles in new situations.
Infer: Buoyancy and Archimedes principle [P2-PH] → The Floating and Sinking Conditions of Objects and Their Application [P2-PH] Realize that the buoyancy and gravity determine the rise and fall of an object.
Apply: Map [P2-GE] → Recognizing Continent [P2-GE] Use maps and other data to describe the geographical location of a continent (such as latitude, hemisphere, and land sea position).
Compare: Photosynthesis (preliminary) [P2-BI] ↔ Respiratory function (preliminary) [P2-BI] Compare respiration and photosynthesis, and summarize the differences and connections between them.
Organize: Law of Universal Gravitation [P3-PH] ↔ Circular Motion [P3-PH] Integrate the laws of universal gravitation and circular motion and obtain expressions for several characteristic quantities in circular motion.
Self-correlation: Rational Number [P2-MA] Master the addition, subtraction, multiplication, division, multiplication and simple mixed operations of rational numbers.

Table 9: More relations in CogNet-KG with the corresponding learning objectives.

The top-4 Communities for P2-MA
Community-1: Geometry Line Segment Definition, Proposition, Theorem Parallelograms Shape Translation and Rotation Special Parallelograms (Rectangle, Diamond, Square) Circle Axial Symmetry
Community-2: Linear equations and their addition and subtraction Linear equations Multiplication and division of equations Factorization Linear quadratic equations Fractions Linear inequalities (systems)
Community-3: Pythagorean theorem Ruler and compass drawing Triangle Isosceles triangle Right angled triangle Congruent triangle
Community-4: Binary linear equations Basic concepts of functions Linear functions Quadratic functions Inverse proportional functions

Table 10: The results of community detection for P2-MA.

The top-4 Communities for P3-PH
Community-1: Electric potential energy Electrostatic phenomenon Law of conservation of charge Coulomb's law Motion (interaction) of charged particles in an electrostatic field Electric field Electric field strength Electrostatic field and its application Electric field lines Comprehensive application of charged particles in electric and magnetic fields Utilization and protection of static electricity power source and current Potential difference Electric potential Magnetic field Magnetic field and its application - Ampere force Magnetic induction strength Magnetic flux Magnetic induction lines Lorentz force Capacitor Motion (energy) of charged particles in an electric field
Community-2: Newton's third law Weightlessness and overweight Comprehensive application of Newton's laws of motion Experiment: Exploring the relationship between acceleration, force, and mass Gravity Elasticity Friction Balance of Common Forces Synthesis and decomposition of forces Newton's first law Newton's second law Centripetal acceleration Mechanical unit system Circular motion Centripetal Force Circular motion in life Planetary motion Universal gravitation law Application of universal gravitation theory Relativity
Community-3: The law of conservation of energy Basic concepts of molecular dynamics Experiment: Estimating the size of oleic acid molecules using oil film method Molecular internal energy Gas experiment law Relationship between gas molecular motion and pressure Second law of thermodynamics Temperature scale Energy Solid Liquid Changes in work, heat, and internal energy First law of thermodynamics
Community-4: Blackbody radiation Particle fluctuation Photoelectric effect Atomic structure model Structure of atomic nucleus Decay of radioactive elements Nuclear fission and fusion Nuclear binding energy

Table 11: The results of community detection for P3-PH.

Top-n	P2-MA	P3-PH
1	Rectangular coordinate system	Comprehensive application of Newton's laws of motion
2	Quadratic function	Conservation of momentum
3	Triangle	Comprehensive application of charged particles in electric and magnetic fields
4	Linear function	Weightlessness and overweight
5	Shape translation and rotation	Newton's second law

Table 12: The top 5 knowledge nodes with the highest betweenness centrality in P2-MA and P3-PH.

MA	PH
The application of trigonometric functions Comprehensive application of unequal relationships Basic Elementary Functions Equation of a straight line	The comprehensive application of Newton's laws of motion Speed and its measurement Utilization and protection of static electricity The conversion and conservation of energy

Table 13: Taking MA and PH as examples of two subjects, we select the relations whose knowledge points belong to different grades and list some samples with high frequency occurrences among them.

Subject Combination	Example
MA & PH	(The concept of vectors, Synthesis and decomposition of forces)
PH & CH	(Molecular internal energy, Thermal effect of chemical reaction)
CH & BI	(Biological macromolecules, Molecular composition of cells)
BI & GE	(Ecosystem, Environmental issues and sustainable development)

Table 14: Taking P3 as the example grade (for there are few cross-subject relations in P2 grade), we select the relations whose knowledge points belong to different subjects and list some samples from different subject combination.

Prompt for CogDialogue-QA Construction
<p>Your task is to design a teacher-student dialogue in a secondary school teaching scenario based on the following content: 1) The current learning objective is a relation which also includes its corresponding knowledge points, cognitive logic relation type. 2) The candidate set for subsequent learning objectives includes multiple learning objectives, which are marked sequentially with numbers.</p> <p>Role Setting: Teacher can help student consolidate his/her knowledge state through two behaviors. 1) Step-1: Explain the current learning objective in one or two sentences, requiring the content to be rigorous, logically clear, highlight key points, and appropriately achieve colloquial, friendly, and in line with the cognitive level of secondary school students. Then generate a related exercise or question to test student's mastery on it, requiring diverse question forms and triggering student's thinking. 2) Step-2: Based on the student's answer, use one or two sentences to assess and summarize his/her mastery on the current learning objective. Then select the most suitable one from the candidate list to realize a natural transition and repeat Step-1 to realize high-quality tutoring. Teacher must not directly quote the original words or texts from the learning objectives in his/her output. Instead, he/she should make significant revisions in terms of vocabulary, sentence structure, and expression style based on their understanding, in order to conform to the style of daily teacher-student dialogue.</p> <p>Student has {mastery level} on the current learning objective. According to the above settings, you need to combine the knowledge ability, cognitive level, and language style of a secondary school student to simulate his/her answers to teacher's question, and try to avoid using meaningless filler words as much as possible to make the dialogue more concise.</p> <p>Current Learning Objective: {current learning objective} Candidate List for Next Learning Objective: {candidate list} Output Format: {"Teacher-ask": teacher_ask, "Student-answer": student_answer, "Choice": index, "Teacher-tutoring": teacher_tutoring }</p>

Table 15: Prompt for CogDialogue-QA dataset construction.

<p>Current Learning Objective: Self-correlation Rational Number [P2-MA] Represent rational numbers with points on the number line and find the opposite and absolute values of rational numbers.</p>
<p>Candidate List for Next Learning Objective:</p> <ol style="list-style-type: none"> 1. Expand (Rational Number [P2-MA] → Real Number [P2-MA]): Recognize and understand that some concepts and operations of rational numbers are still applicable in real numbers. 1. Expand (Rational Number [P2-MA] → Real Number [P2-MA]): Recognize and understand that some concepts and operations of rational numbers are still applicable in real numbers. 2. Classify (Rational Number [P2-MA] → Real Number [P2-MA]): Describe the relationship between integers, fractions, rational numbers, irrational numbers, real numbers, etc. 3. Expand (Rational Number [P2-MA] → Real Number [P2-MA]): Search for patterns and make reasonable inferences based on computational rules. 4. Apply (Rational Number [P2-MA] → Quadratic Radical [P2-MA]): estimate the approximate range of a quadratic root using rational numbers. <p>(Omit the rest.)</p>
<p>Teacher-ask: Today we will learn how to represent rational numbers using a number line, while also mastering the method of finding the opposite number and absolute value of a number. The number line is a horizontal straight line that clearly represents the position of rational numbers. For example, positive numbers to the right, negative numbers to the left, and 0 in the middle. So the question is: if the rational number -5 corresponds to point P on the number axis, which rational number is its opposite? How to express its absolute value?.</p>
<p>Student-answer: The opposite of rational number -5 is 5, which can be represented on the number axis as point Q symmetric about the origin with respect to point P. Its absolute value is 5, corresponding to two points on the number axis that are 5 degrees away from the origin, namely -5 and 5.</p>
<p>Choice: 2</p>
<p>Teacher-tutoring: Excellent! You are now proficient in representing rational numbers and finding their inverse and absolute values. Next, we will further explore and consider the relationship between rational numbers and real numbers. How do integers, fractions, rational numbers, irrational numbers, and real numbers contain each other? Which set is the largest?</p>

Table 16: An example for the CogDialogue-QA dataset. The value of "choice" demonstrates that the teacher provides the following tutoring based on the second learning objective in the candidate list.

Scoring Dimensions and Criteria for Instructional Content
<p>Clarity: Evaluates the legibility, specificity, and appropriateness of teaching objectives and content expression.</p> <p>1 point: The teaching objectives reflected in the dialogue are vague or missing, the content is obscure and difficult to understand, and seriously deviates from the students' cognitive level.</p> <p>2 points: The teaching objectives reflected in the dialogue are relatively broad, and there are many difficult to understand aspects in the content expression, which is generally matched with the learning situation.</p> <p>3 points: The teaching objectives reflected in the dialogue are basically clear, the content is generally understandable, and generally in line with the students' current cognitive stage.</p> <p>4 points: The teaching objectives reflected in the dialogue are clear and specific, the content is well organized, and easy for students to understand and accept.</p> <p>5 points: The teaching objectives reflected in the dialogue are highly clear, specific, and measurable, and the content is presented accurately to match the students' cognitive level, which is highly inspiring.</p>
<p>Expression: Evaluates the organization logic and matching degree of teaching language with students' cognition and knowledge proficiency.</p> <p>1 point: Expression is chaotic, illogical; language is overly academic or childish, making it difficult to understand.</p> <p>2 points: Poor organization; language is noticeably disconnected from students' cognitive level.</p> <p>3 points: Expression is basically organized; language roughly matches students' comprehension level.</p> <p>4 points: Language is clear, logical, and flexibly adjusted to align with student cognition.</p> <p>5 points: Language organization is highly logical and inspiring; wording is precise and closely matches students' cognitive and linguistic style; expression is vivid.</p>
<p>Coherence: Evaluates the naturalness and logical flow of contextual memory and connection within the dialogue.</p> <p>1 point: Dialogue is fragmented, with little reference to context; transitions are stiff or broken.</p> <p>2 points: Can only connect to the immediately previous turn; overall dialogue flow is unclear.</p> <p>3 points: Can remember main recent content, maintaining basic coherence.</p> <p>4 points: Can effectively use dialogue history for reasonable connections and extensions; conversation flows smoothly.</p> <p>5 points: Dialogue is highly coherent, naturally weaving scattered information into a structured, in-depth sustained conversation.</p>
<p>Constructiveness: Evaluates the degree to which the logical progression and cognitive connection between knowledge points across the dialogue facilitate systematic, structured knowledge building.</p> <p>1 point: Knowledge fragments are disjointed, with jumps or repetition; fails to promote integration.</p> <p>2 points: Knowledge is barely linked, with unclear progression; organization is scattered.</p> <p>3 points: Knowledge has some logical connections, supporting preliminary integration.</p> <p>4 points: Knowledge content is logically clear and progressively ordered, helping students build connections.</p> <p>5 points: Knowledge is highly structured and hierarchical, naturally promoting deep cognitive integration and knowledge system construction.</p>

Table 17: The 5-point scale for the tutorial interaction quality evaluation system (instructional content).

Scoring Dimensions and Criteria for Interactive Support
<p>Diagnosis: Evaluates the accuracy in identifying students' knowledge state, core misunderstandings, and cognitive obstacles.</p> <p>1 point: Completely fails to recognize knowledge points or errors in student responses; no diagnostic behavior.</p> <p>2 points: Can only identify surface-level correctness, struggling to pinpoint specific weak points or cognitive obstacles.</p> <p>3 points: Can identify core knowledge points and general grasp level, making preliminary judgments about weaknesses.</p> <p>4 points: Accurately identifies mastery level and can diagnose specific issues like concept confusion or logical gaps.</p> <p>5 points: Deeply and precisely diagnoses knowledge mastery, thought processes, and cognitive obstacles, providing a solid foundation for personalized support.</p>
<p>Guidance: Evaluates the ability to stimulate thinking, promote reflection, and support autonomous learning planning through questioning or prompting.</p> <p>1 point: Provides answers or instructions directly; lacks guiding questions or reflective support.</p> <p>2 points: Guidance is rigid and mechanical, failing to promote deep thinking or self-monitoring.</p> <p>3 points: Can ask basic follow-up questions based on student responses, offering some reflection or task-planning prompts.</p> <p>4 points: Uses step-by-step questioning to effectively guide self-correction and deep thinking, supporting goal-setting and planning.</p> <p>5 points: Questioning is masterful, naturally guiding knowledge construction and reflection on learning strategies, effectively fostering meta-cognition and self-directed learning.</p>
<p>Feedback: Evaluates the accuracy, constructiveness, and specificity of explanations, evaluations, and improvement suggestions in response to errors.</p> <p>1 point: Only indicates right/wrong, or provides no feedback.</p> <p>2 points: Feedback is vague, failing to explain error causes or provide effective improvement suggestions.</p> <p>3 points: Can point out error causes and provide basic explanations and conventional suggestions.</p> <p>4 points: Accurately analyzes root causes, explains clearly, and gives specific, actionable suggestions.</p> <p>5 points: Provides profound, thorough attribution analysis; evaluation is constructive; improvement suggestions are specific, personalized, and motivating for growth.</p>

Table 18: The 5-point scale for the tutorial interaction quality evaluation system (interactive support).

Scoring Dimensions and Criteria for Individual Development
<p>Personalization: Evaluates the ability to tailor learning paths and content based on student readiness, cognitive state or style.</p> <p>1 point: Completely standardized feedback, ignoring individual differences.</p> <p>2 points: Only superficial personalization (e.g., using name); learning path is rigid.</p> <p>3 points: Can provide slightly different examples or suggestions based on explicit knowledge level.</p> <p>4 points: Can combine knowledge level and cognitive state to offer noticeably differentiated learning paths or strategy suggestions.</p> <p>5 points: Deeply integrates knowledge level, cognitive traits, learning style, and even emotional state to provide highly customized, dynamically adjusted learning paths and support.</p>
<p>Motivation: Evaluates the effectiveness in boosting student confidence and intrinsic motivation through emotional interaction and attribution style.</p> <p>1 point: Language is indifferent, only negative, or ignores student emotional needs.</p> <p>2 points: Motivation is superficial, lacking genuine feeling or concrete basis.</p> <p>3 points: Can offer basic affirmation and encouragement, attempting to create a positive atmosphere.</p> <p>4 points: Can sincerely acknowledge effort and progress, emphasize growth, effectively enhancing student confidence and engagement.</p> <p>5 points: Skillfully employs growth mindset, empathy, and positive vision to significantly ignite intrinsic learning motivation and resilience.</p>

Table 19: The 5-point scale for the tutorial interaction quality evaluation system (individual development).