Multi-Perspective Consolidation Enhanced Cognitive Diagnosis via Conditional Diffusion Model

Guanhao Zhao[1,2], Zhenya Huang[1,2], Cheng Cheng[2], Yan Zhuang[2], Qingyang Mao[1,2], Xin Li[2,3], Shijin Wang[2,3*], Enhong Chen[1,2]

[1] The School of Artificial Intelligence and Data Science, University of Science & Technology of China, Hefei, China
[2] State Key Laboratory of Cognitive Intelligence, Hefei, China
[3] IFLYTEK Research, Hefei, China
{ghzhao0223, doublecheng, zykb, maoqy0503}@mail.ustc.edu.cn,
{huangzhy, leexin, cheneh}@ustc.edu.cn, sjwang3@iflytek.com

Abstract

Cognitive diagnosis, which assesses the learners' competence from learners' interaction logs, plays a vital role in education. It provides a crucial reference for gauging learners' proficiency levels and tailoring future learning activities accordingly. Researchers have proposed numerous cognitive diagnosis models to address this task. Despite their success, these models continue to face the ill-posed problem because of the information loss caused by under-expressive interaction function and incomplete observations. In this paper, we address these challenges by proposing a novel cognitive diagnosis model, DMC-CDM, based on the theoretical premise that cognitive states can be captured with minimal information loss by maximizing the mutual information between observed and potential observations. Specifically, DMC-CDM incorporates a semantic extractor to provide a comprehensive semantic understanding of learners' interaction logs, thereby enhancing current collaborative-based cognitive state representations. It then consolidates multi-perspective observations to capture precise cognitive states by maximizing mutual information between these observations. We conducted extensive experiments on three datasets, and the experimental results demonstrate that our proposed model is both effective and beneficial for downstream applications in education.

1 Introduction

Accurate assessment of cognitive states is crucial in educational applications. It provides critical insights for both learners and educators to comprehend learners' current cognitive levels and modify learning paths appropriately (Jiang et al. 2023). This assessment also plays a vital role in evaluating and ranking learners, which is essential for merit-based selections like the GRE and GMAT (Ghosh and Lan 2021). As shown in Figure 1, cognitive diagnosis (Wang et al. 2023) leverages learners' personal practice logs to estimate their cognitive states. This process involves recovering latent signals (i.e., cognitive states) from observed data (i.e., interactions), and characterizing it as a typical inverse problem.

Numerous cognitive diagnosis models (CDMs) have been developed to capture learners' cognitive states (Lord 1980; De La Torre 2009; Zhao et al. 2024). These methods typically start by designing an interaction function based on prior knowledge, mapping latent cognitive states to interaction logs. Learners' cognitive states are then captured through parameter estimation. Despite their success, these methods often encounter a critical issue: the cognitive diagnosis process is inherently ill-posed due to information loss (Song et al. 2022). This means that given an observation (e.g., learner's interaction logs in Exam 3), it is challenging to precisely recover the signals (i.e., cognitive states). The information loss typically stems from two main causes: (1) *Under-expressive Interaction Function:* Previous models heavily rely on manually designed interaction functions to map the cognitive states to interactions and generally overlook richer information within the logs, leading to oversimplified interaction functions and information loss during diagnosis, particularly when logs are sparse (Dai et al. 2021; Zhao et al. 2023b,a). (2) *Incomplete Observations:* The observed interaction logs are often incomplete; for example, a learner might only answer a few exercises (e.g., the perspective of Exam 3), leading to many unobserved interactions in other perspectives such as Exam 1 and Exam 2, as depicted in the dotted part of Figure 1. Abundant information in the unobserved interactions is lost. Recent advances aim to address the ill-posed nature of cognitive diagnosis by employing more expressive interaction functions (Wang et al. 2023; Gao et al. 2021) or by incorporating unobserved interactions through sampling (Yao et al. 2023; Ma et al. 2024). However, their performance remains limited, and their underlying motivation remains intuitive.

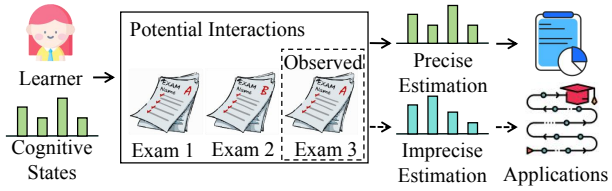

Figure 1: The illustration of cognitive diagnosis and its inherently ill-posed nature, where the dotted line represents the cognitive diagnosis process with partial observations.

In this paper, we uniformly tackle the aforementioned

*Corresponding author

Under-expressive Interaction Function and *Incomplete Observations* problem by a theoretically grounded approach, which is achieved by precisely capturing cognitive states from a single perspective (e.g., Exam 3) and maximizing the mutual information between perspectives (e.g., Exam 3 and Exam 2), thus consolidating the cognitive states and enhancing the effectiveness. We implement the theoretical basis with a novel **D**iffusion-based **M**ulti-perspective **C**onsolidated **C**ognitive **D**iagnosis **M**odel, designated as DMC-CDM. Specifically, for a more expressive interaction function, we introduce a semantic extractor to extract semantic information from interaction logs (Ren et al. 2024) and combine the semantic information with the collaborative information using a cross-attention strategy, capturing more information from a single perspective. Furthermore, we consolidate the captured learners' cognitive states through a conditional diffusion-based multi-perspective mutual information maximization scheme. We introduce noise to an observed perspective (e.g. Exam 3) and subsequently reconstruct it by conditioning on observations from other perspectives (e.g., Exam 1), further maximizing the mutual information between them and consolidating information from different perspectives.

In summary, the main contributions of this paper can be delineated as follows: (1) To the best of our knowledge, this paper is the first to address the ill-posed problem in cognitive diagnosis by leveraging a theoretically grounded framework for multi-perspective consolidation. (2) The proposed DMC-CDM represents the first model that captures cognitive states from a single perspective precisely and consolidates the multi-perspective cognitive states. (3) Comprehensive experiments in three real-world datasets provide compelling evidence of the effectiveness of our proposed model and its constituent components.

2 Related Work

Cognitive Diagnosis. Cognitive diagnosis Models (CDMs), which provide cognitive diagnostic reports, help better understand learners' abilities and optimize subsequent learning activities (Jin, Huang, and Wen 2015; Liu et al. 2024; Wang et al. 2024). The dominant approach in CDMs revolves around formulating an interaction function F that links latent cognitive states to learners' responses. The early Item Response Theory-based CDM (IRT) (Lord 1980) utilizes a logistic function to model the interaction function, and Deterministic Inputs, Noisy-And gate (DINA) (De La Torre 2009) models the interaction of learners and exercises with a discrete-continuous product function. Facing the ill-posed problem, researchers have introduced expressive artificial neural networks into cognitive diagnosis, like NeuralCDM (Wang et al. 2023), and DisenCD (Chen et al. 2024b). Moreover, some recent advancements like EIRS (Yao et al. 2023) and CMES (Ma et al. 2024) exploit unobserved interactions by sampling. However, despite their success, these methods do not explore the inherent ill-posed property of cognitive diagnosis, and their motivation is somewhat intuitive. To tackle the information loss problem, we propose a theoretically grounded model that is designed for precise capture from a single perspective and effective consolidation from multiple perspectives.

Solving Inverse Problem via Diffusion Model. The inverse problem, focused on recovering underlying signals from observations, has been a significant topic in academic research for a long time (Song et al. 2022). Traditionally, approaches to solving inverse problems depend on gathering a large quantity of both observations and latent signals (Shen, Zhao, and Xing 2019). In recent years, the advent of the Diffusion Model has precipitated a notable shift in this approach. Numerous scholars have adopted the Diffusion Model to address inverse problems, owing to its demonstrated advantages over traditional models in modeling data distribution (Song et al. 2022; Xiao et al. 2024; Nichol and Dhariwal 2021; Chung et al. 2022). Despite their success, these methods cannot be directly applied to cognitive diagnosis, as they require modeling the distribution of latent signals (i.e., cognitive states), which are not readily accessible in cognitive diagnosis scenarios. In this paper, we adopt the same assumption made in (Tewari et al. 2024) to recover cognitive states from all potential observations. However, our approach goes beyond that of (Tewari et al. 2024) by offering a comprehensive theoretical analysis of the consolidation process in cognitive diagnosis, based on mutual information.

3 Methodology

3.1 Problem Setup

Let $\mathcal{S} = \{s_1, s_2, \ldots, s_{|\mathcal{S}|}\}$ represents the set of learners, and $\mathcal{Q} = \{q_1, q_2, \ldots, q_{|\mathcal{Q}|}\}$ denotes the set of exercises within an educational system. Each learner interacts with the system by engaging with a selection of these exercises. The total interaction logs are structured as a set of triplets, denoted by $R = \{(s, q, r)|s \in \mathcal{S}, q \in \mathcal{Q}, r \in [1, 0]\}$, where $r_{i,j}$ indicates whether learner s_i answered exercise q_j correctly or not (i.e., 1 denotes the correct answer and 0 denotes the wrong answer). The latent cognitive states of the learners are represented by a set $\mathbf{\Theta} = \{\theta_i|s_i \in \mathcal{S}\}$. Concurrently, the latent features of the exercises are encapsulated by $\mathbf{B} = \{\beta_j|q_j \in \mathcal{Q}\}$.

Given the learners' interactions logs R, the objective of Cognitive Diagnosis is to estimate learners' latent cognitive states Θ precisely by maximizing:

$$p(\mathbf{\Theta}|R) \propto p(R|\mathbf{\Theta}, \mathbf{B})p(\mathbf{\Theta}), \tag{1}$$

where $p(\mathbf{\Theta})$ is the prior of cognitive states $\mathbf{\Theta}$, and $p(R|\mathbf{\Theta}, \mathbf{B})$ is modeled by the interaction function $F : \mathbf{\Theta} \times \mathbf{B} \rightarrow \mathbb{R}$ in previous cognitive diagnosis models. For example, for learner s_i and exercise q_j, the performance is $\hat{r}_{i,j} = F(\theta_i, \beta_j)$. It is noteworthy that during the parameter estimation process for cognitive states $\mathbf{\Theta}$, the exercise features $\mathbf{B}$ are typically held constant, having been either manually configured or initially learned (Wang et al. 2023).

3.2 Theoretical Basis

In section 1, we have concluded the reasons for information loss are *Under-expressive Interaction Function* and *Incomplete Observations*, which render cognitive diagnosis an ill-posed inverse problem. In this paper, we propose a theoretically grounded framework to solve the ill-posed cognitive diagnosis problem, as illustrated in Figure 2(a). To begin

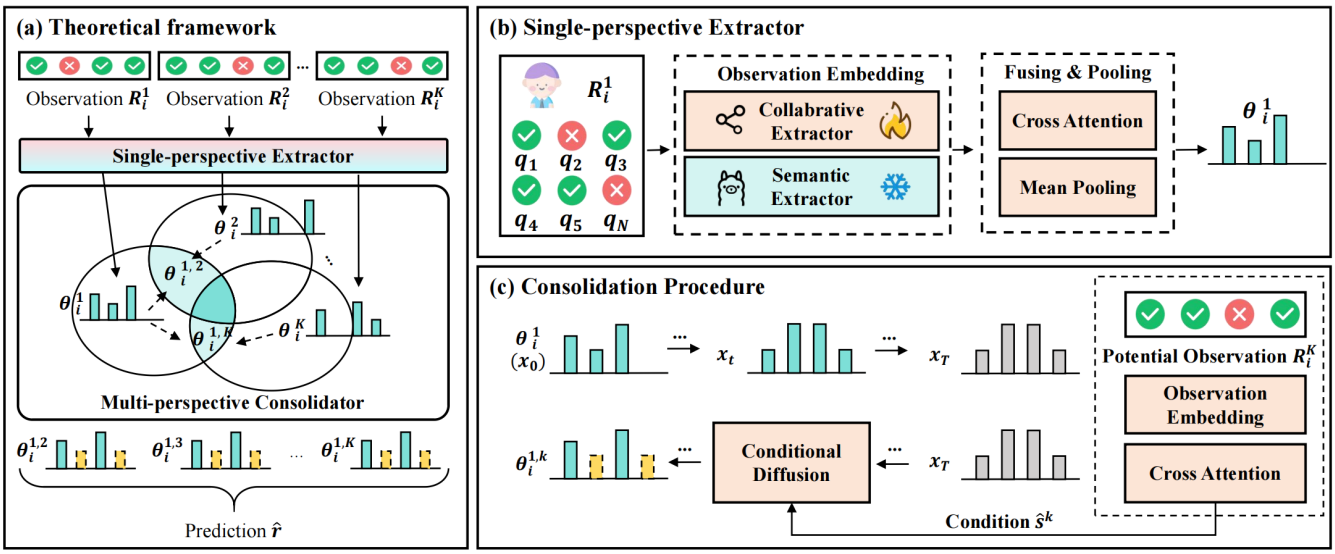

Figure 2: (a) The overall framework of the proposed DMC-CDM, (b) Single-perspective Extractor, and (c) The consolidation procedure of two perspectives with conditional diffusion.

with, we adopt a assumption (Tewari et al. 2024), which is simple but rational.

Assumption 1. *The latent cognitive states can be recovered with all possible interaction observations.*

This assumption is grounded in common sense. For instance, each teacher has a unique style of designing exams, which can lead to bias when assessing students through a single exam. However, by aggregating results from multi-perspective of exams, we can achieve a more accurate assessment of a student's cognitive state.

Based on the assumption, we denote the observed interaction logs as R^1 (i.e., Exam 3 in Figure 1) and other possible interaction logs as R^k (i.e., Exam 1 and 2 in Figure 1), our goal can be described as follows:

$$\max \mathbb{E}_{p(\cdot)}[p(\mathbf{\Theta}|R^1, R^2, ..., R^K)], \tag{2}$$

where $p(\cdot)$ denotes the joint probability distribution of $R^1, R^2, ...R^K$. One step further, R^1 contains partial information from $\mathbf{\Theta}$ and so do $R^2, ..., R^K$. The shared information (i.e., The intersection part of the Venn diagram in Figure 2(a)) between them would be the actual information needed for capturing $\mathbf{\Theta}$. Since R^1 is known, the above equation can be reformulated as utilizing information underlying $R^2, ..., R^K$ to maximize such a posterior:

$$\max \mathbb{E}_{p(\cdot)}[p(\mathbf{\Theta}, R^1|R^2, ..., R^K)], \tag{3}$$

Theorem 1. *Maximizing the posterior in Equation (3) to reconstruct the hidden latent cognitive states Θ can be approximated by maximizing the mutual information between the observed R^1 and other possible observations $R^2, R^3, \ldots, R^K$.*

Drawing from Theorem 1, we establish a theoretical basis for addressing the ill-posed cognitive diagnosis problem by integrating information from other potential observations R^k into the observed R^1. Building on the above theoretical basis, the ill-posed cognitive diagnosis problem would be alleviated. The proof of Theorem 1 is as follows:

Proof. The observed R^1 is fixed.

$$\mathbb{E}_{p(\cdot)}[p(\mathbf{\Theta}, R^1|R^2, ..., R^K)] \tag{4}$$

$$\propto \mathbb{E}_{p(\cdot)} \log[\int_{\mathbf{\Theta}} \frac{p(\mathbf{\Theta}, R^1|R^2, ..., R^K)}{p(R^1)} d\mathbf{\Theta}] \tag{5}$$

$$= \mathbb{E}_{p(\cdot)} \log[\frac{p(R^1|, R^2, ..., R^K)}{p(R^1)}] \tag{6}$$

$$= \mathbb{E}_{p(\cdot)} \log[\frac{p(R^1, R^2, ..., R^K)}{p(R^1)p(R^2, ..., R^K)}] \tag{7}$$

$$= I(R^1; (R^2, ..., R^K)) \tag{8}$$

$$\leq I(R^1; R^2) + I(R^1; R^3) + ... + I(R^1; R^K). \tag{9}$$

□

In practice, interaction logs (i.e., observations) are discrete, making it challenging to model the mutual information directly. To address this, we implement our theoretical framework by focusing on consolidating cognitive states instead of the observations themselves. Hence, the remaining issues for implementation are: (1) How can learners' cognitive states be captured more precisely from a single perspective? and (2) How to effectively consolidate the captured cognitive states from different perspectives.

3.3 Model Implementation

Our approach comprises two main components: the Single-perspective Extractor (SPE) and the Multi-perspective Consolidator (MPC). The SPE focuses on extracting latent cognitive states by integrating observed interactions from collaborative and semantic viewpoints (Ren et al. 2024), further enhancing the expression of the interaction function

F. Subsequently, the MPC employs a conditional diffusion model (Peebles and Xie 2023) to cover the manually noised cognitive states from one perspective based on specific perspectives, thereby optimizing the mutual information across different observational perspectives. Ultimately, we integrate the predicted performance from various perspectives using a weighted average method.

Single-perspective Extractor Previous Cognitive Diagnosis models, such as those documented in (Wang et al. 2023; Gao et al. 2021, 2023a), predominantly utilize collaborative information to derive learners' latent cognitive states. However, these models often encounter limitations when the available collaborative data is sparse (Dai et al. 2021). In response, we have integrated an semantic extractor with a collaborative extractor to process observed learner interactions (Ren et al. 2024). These are combined using a cross-attention mechanism, as depicted in Figure 2(b).

Observation Embedding: Given the interaction logs $[(q_1, r_{i,1}), (q_2, r_{i,2})..., (q_N, r_{i,N})]$ of learner s_i, we process the sequence of exercises $[q_j]$ from R_i through two distinct embedding transformations, i.e., collabrative extractor and semantic extractor (Ren et al. 2024). Based on this, we can map the observation R_i as semantic representations $\mathbf{S}^S = [\mathbf{q}_j^S] \in \mathbb{R}^{N \times d_1}$ and collaborative representations $\mathbf{S}^C = [\mathbf{q}_j^C] \in \mathbb{R}^{N \times d_1}$, where N the length of the learner s_i's interaction logs.

Fusing & Pooling: We then employ a cross-attention mechanism to enhance the alignment between the information contained within the semantic representations $\mathbf{S}^S$ and collaborative representations $\mathbf{S}^C$, further fusing the information. Following this, we concatenate $\hat{\mathbf{S}}^S$ and $\hat{\mathbf{S}}^C$ to produce $\hat{\mathbf{S}} = [\hat{\mathbf{q}}_j] \in \mathbb{R}^{N \times 2d_1}$. One step further, we utilize the mean pooling operation to get learner s_i's hidden cognitive states $\theta_i \in \mathbb{R}^{2d_1}$.

Given learner's cognitive states θ_i and exercises' features $\hat{\mathbf{q}}_j$. We can predict the performance $\hat{r}_{i,j} = F(\theta_i, \hat{\mathbf{q}}_j)$ like previous CDMs (Wang et al. 2023). Furthermore, the SPE is trained with the cross-entropy loss:

$$\mathcal{L}_{ce} = \mathbb{E}[r \log(\hat{r}) + (1 - r) \log(1 - \hat{r})], \tag{10}$$

where r and $\hat{r}$ denote real and predicted performance.

Multi-perspective Consolidator After addressing the *under-expressive interaction function* with SPE, it is crucial to recognize the issue of *incomplete observations*. This limitation primarily arises because learners may engage with only a subset of exercises (e.g., Exam 3 in Figure 1), leaving many interactions unrecorded (e.g., Exam 1 and 2 in Figure 1). These gaps in data can significantly hinder the accuracy of cognitive diagnosis. Existing methods (Yao et al. 2023; Ma et al. 2024) utilize a sampling method to identify reliable unobserved interactions as pseudo-labels. However, their methodology, while intuitive, lacks a solid theoretical foundation. Fortunately, our theoretical analysis provides an effective solution by maximizing the mutual information between observed interactions and potential unobserved interactions. To implement this, we introduce the Multi-perspective Consolidator, which amalgamates information from various perspectives. The workflow is outlined as follows:

Possible Observations Retrieval: While it would be ideal to access all potential observations $R^1, R^2, ..., R^K$, in practice, we are only able to get a fraction of them (e.g., R^1). Thus, we propose a retrieval method to get potential observations from similar learners, thereby augmenting the potential observations. Without loss of generalization, for a learner s_i, with its single-perspective cognitive states θ_i, we can retrieve the top K most similar learners with cosine similarity. After retrieving learners, we denote them as $[\theta_\mathbf{i}^\mathbf{1}, \theta_\mathbf{i}^\mathbf{2}, \theta_\mathbf{i}^\mathbf{3}, ..., \theta_\mathbf{i}^\mathbf{K}]$ and their corresponding interaction logs as $[R_i^1, R_i^2, R_i^3, ...R_i^K]$.

Mutual Information Maximum: After retrieving and augmenting potential observations, we can consolidate the multi-perspective observations by maximizing the mutual information between the observed R_i^1 and potential observations $[R_i^2, ..., R_i^K]$. Given that this paper focuses on cognitive states, we aim to maximize the mutual information between cognitive states observed from different perspectives. We achieve this by adopting a conditional diffusion approach as outlined in (Ho, Jain, and Abbeel 2020; Peebles and Xie 2023; Yang et al. 2024b). We perturb the observed cognitive states $\theta_\mathbf{i}^\mathbf{1}$ (i.e., x_0) and subsequently recover it using the information underlying the potential observation R_i^k, as depicted in Figure 2(c). During the recovery process, we leverage the information from R_i^k to reconstruct $\theta_\mathbf{i}^\mathbf{1}$ initially learned from R_i^1, thus effectively maximizing the mutual information $I(R_i^1; R_i^k)$ to a certain degree.

Specifically, we first gradually introduce Gaussian noise ϵ in x_0 (i.e., the captured cognitive states $\theta_\mathbf{i}^\mathbf{1}$ from a single-perspective observation R_i^1):

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{\alpha_t}\mathbf{x}_{t-1}, (1 - \alpha_t)\mathbf{I}), \tag{11}$$

where α_t is the variance schedule, $\mathcal{N}$ denotes the Gaussian distribution. Due to the additivity of Gaussian process, let $\epsilon \in \mathcal{N}(0, 1)$ and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$, we can directly get $\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon$. Next, we encode the potential observation R_i^k the same as the SPE to obtain the representation $\hat{\mathbf{S}}_i^k \in \mathbb{R}^{N \times 2d_1}$, which further serves as the condition for the recovery process:

$$p_\omega(\mathbf{x}_{t-1}|\mathbf{x}_t, \hat{\mathbf{S}}^\mathbf{k}) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_\omega(\mathbf{x}_t, \hat{\mathbf{S}}_i^k, t), \Sigma_\omega(\mathbf{x}_t, \hat{\mathbf{S}}_i^k, t)). \tag{12}$$

Currently, we have defined the form of the noising and recovering process, which aims to extract helpful information from the condition (i.e., potential observation R_i^k) to recover the target $\theta_\mathbf{i}^\mathbf{1}$, thereby further maximizing the mutual information between them.

Moreover, we will introduce how to optimize the recovery process. In this paper, following (Yang et al. 2024b), we define the reparameterization of $\mu_\omega(\mathbf{x}_t, \hat{\mathbf{S}}_i^k, t)$ to predict $\mathbf{x_0}$ rather than the added noise ϵ:

$$\mu_\omega(\mathbf{x}_t, \hat{\mathbf{S}}_i^k, t) = \frac{\sqrt{\alpha_t}(1 - \bar{a}_{t-1})}{1 - \bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}b_t}{1 - \bar{\alpha}_t} D_\omega(\mathbf{x}_t, \hat{\mathbf{S}}_i^k, t), \tag{13}$$

where $D_\omega(\mathbf{x}_t, \hat{\mathbf{S}}_i^k, t)$ represents the prediction of x_0 (i.e., $\theta_\mathbf{i}^{\mathbf{1},\mathbf{k}}$), which is consolidated from R_i^1 and R_i^k. The architec-

ture of D_ω is implemented by Diffusion Transformer (Peebles and Xie 2023) shown in Figure 2(c). Based on the reparameterization form, our learning objective is:

$$\mathcal{L}_{Diff} = \mathbb{E}_{t,\mathbf{x_0}}[\|D_\omega(\mathbf{x}_t, \hat{\mathbf{S}}_i^k, t) - \mathbf{x_0}\|^2]. \quad (14)$$

Due to space constraints and the focus of this paper, we will not elaborate on the score matching loss $\mathcal{L}_{Diff}$ here. For a detailed derivation of it, interested readers are referred to (Song et al. 2020; Yang et al. 2024b) for comprehensive explanations. During the training process, we randomly mask $\hat{\mathbf{S}}_i^k$ as 0 to ensure the condition is learned (Ho and Salimans 2022; Peebles and Xie 2023). Upon obtaining the learned $D_\omega(\mathbf{x}_t, \hat{\mathbf{S}}_i^k, t)$, we can systematically recover the consolidated $\theta_i^{1,k}$, which reconciles the two perspectives (R_i^1 and R_i^k), by following the stepwise process outlined in Equation (12). Finally, to ensure that the consolidated $\theta_i^{1,k}$ effectively incorporates information from both R_i^1 and R_i^k, thereby maximizing the mutual information, we utilize these consolidated cognitive states to predict the performance on exercise lists in R_i^1 and retrieved R_i^k, further minimizing the reconstruction loss, denoted as $\mathcal{L}_{Recon}$. The computation of this loss is according to Equation (10).

The training process for our proposed DMC-CDM model comprises two primary phases: Initially, we train the SPE using $\mathcal{L}_{ce}$, focusing on recognizing cognitive states from a singular perspective. Subsequently, we utilize both $\mathcal{L}_{Diff}$ and $\mathcal{L}_{Recon}$ to refine the training of the SPE and MPC.

Model Inference Precisely capturing θ_i^1 from a single perspective (e.g., Exam 3) and maximizing the mutual information between observed and potential observations (e.g., Exam 1 and 2) through a noising and conditional recovery process, we obtain the consolidated cognitive states $[\theta_i^{1,k}, \theta_i^{1,2}, ..., \theta_i^{1,k}]$. The final predicted performance of the learner s_i on a typical exercise q_j is computed using a weighted sum approach:

$$\hat{r}_{i,j} = \sum_k^K W_{cos}^{1,k} F(\theta_i^{1,k}, \mathbf{q_j}), \quad (15)$$

where $W_{cos}^{1,k}$ represents the normalized cosine similarity between θ_i^1 and $\theta_\mathbf{i}^\mathbf{k}$, calculated during the retrieval of potential observations. And the interaction function F is implemented according to (Wang et al. 2023).

4 Experiments

We have designed a series of detailed experiments to explore the following research questions: (**RQ1**) How prominent is DMC-CDM in the learners' performance prediction? (**RQ2**) What role does each component of the DMC-CDM play in its overall performance? (**RQ3**) How does the DMC-CDM perform in different scenarios? (**RQ4**) How does our model provide value in real-world applications?

4.1 Datasets

We conduct experiments on three widely used educational benchmark datasets: **MoocRadar**[1], an open dataset that provides interaction logs with responses and knowledge concepts; **Ifly**[2], a dataset collected from the iFLYTEK learning machine that includes interaction logs on mathematics; and **Junyi**[3], a large-scale learning interaction dataset collected from the Junyi educational platform.

Dataset	MoocRadar	Ifly	Junyi
Learners	3,445	21,068	21,902
Exercises	918	9,653	586
Concepts	71	1,371	586
Interactions	159,531	516,613	1,072,267
Avg. Length	46.3	24.5	49.0
Sparsity	5.04%	0.254%	8.35%

Table 1: Dataset statistics.

To preprocess the data, we first filter the questions and knowledge concepts that the number of interactions is less than 30. Then, we filter learners whose interactions are less than 30. Table 1 shows the basic statistics of the datasets after preprocessing. We randomly split each learner's interaction log with the proportion $80\% : 10\% : 10\%$ to get *train set*, *validation set*, and *test set*, where the *train set* is used to train our proposed model, *validation set* is used to tune hyper-parameters, and *test set* is used to test.

4.2 Experimental Setup

Baselines We compare the performance of the DMC-CDM with that of commonly used CDMs. These baselines are described below: (1) IRT (Lord 1980) utilizes a logistic-like interaction function to model scalar learners' cognitive states and their performance. MIRT (Ackerman, Gierl, and Walker 2003) extends IRT by modeling learners' cognitive states with multidimensional vectors. NeuralCDM (Wang et al. 2023) utilizes the neural networks to capture learners' cognitive states from interactions. (2) NeuralCDM+ (Wang et al. 2023) further incorporate text information with word2vec embeddings, KaNCD (Wang et al. 2023) extends NeuralCDM with knowledge-association based information. RCD (Gao et al. 2021) and TechCD (Gao et al. 2023b) introduce a graph-based method to capture collaborative information underlying interactions and further fuse them to capture abilities. (3) EIRS (Yao et al. 2023) and CMES (Ma et al. 2024) exploit unobserved interactions based on sampling mechanism and pseudo-labels, further capturing learners' abilities.

Evaluation Protocol To effectively assess Cognitive Diagnosis Models (CDMs), researchers commonly rely on indirect measures due to the impracticality of directly observing the latent cognitive states of learners. Performance prediction tasks on test sets serve as a suitable proxy for this evaluation (Wang et al. 2023; Gao et al. 2021). We selected three metrics for our analysis: the area under the curve (AUC), accuracy (ACC), and F1 score. Higher values in these metrics signify superior performance of the methods under consideration.

[1]https://github.com/THU-KEG/MOOC-Radar

[2]https://xxj.xunfei.cn/

[3]https://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=1198

Model	MoocRadar			Junyi			Ifly		
	AUC ↑	ACC ↑	F1 ↑	AUC ↑	ACC ↑	F1 ↑	AUC ↑	ACC ↑	F1 ↑
IRT	0.6122	0.8699	0.9194	0.8217	0.8246	0.8955	0.7860	0.7112	0.7721
MIRT	0.7352	0.8287	0.9012	0.8171	0.8245	0.8909	0.7044	0.6524	0.6870
NeuralCDM	0.8817	0.8949	0.9409	0.8256	0.8257	0.8967	0.8090	0.7349	0.7719
NeuralCDM+	0.8959	0.8731	0.9323	0.8826	0.8448	0.9092	0.8447	0.7514	0.7798
RCD	**0.9241**	0.9126	0.9502	0.8262	0.7716	0.8835	0.8356	0.7540	0.7912
KaNCD	0.8789	0.8855	0.9342	0.8531	0.8307	0.8939	0.8299	0.7516	0.7920
TechCD	0.7106	0.8741	0.9328	0.8294	0.8142	0.8833	0.8341	0.7546	0.7963
EIRS	0.9108	0.9048	0.9478	0.8281	0.8330	0.8990	0.7989	0.7258	0.7679
CMES	0.9110	0.8988	0.9500	0.8671	0.8184	0.8808	0.8436	0.7590	0.7989
DMC-CDM	0.9064	**0.9143***	**0.9521***	**0.8834**	**0.8597***	**0.9168***	**0.8500***	**0.7614***	**0.8079***
DMC-CDM-w/o M	0.9069	0.9123	0.9506	0.8574	0.8451	0.9100	0.7679	0.6745	0.7621
DMC-CDM-w/o S	0.9010	0.9116	0.9508	0.8626	0.8537	0.9141	0.8097	0.7301	0.7887
DMC-CDM-w/o Both	0.9076	0.9105	0.9503	0.8291	0.8462	0.9104	0.7146	0.6462	0.7451

Table 2: Performance prediction result. All results are obtained by computing the average result of 5 random seeds. And * denotes significant improvement with p-value < 0.05 in the t-test.

Setup Details The *DMC-CDM* and all baselines are implemented by PyTorch. The parameters are all initialized with zero weight in the same way. The dropout rate is set as 0.1. The hidden dim of each neural network is set as 512. We use the Adam (Kingma and Ba 2014) optimizer to optimize our model. The learning rate of the denoise module is 0.0005. The learning rate of the encoders and interaction function is set as 0.0005. The semantic extractor chosen in our experiments is BGE-M3 (Chen et al. 2024a). All experiments are run on a Linux server with four 2.30 GHz Intel Xeon CPUs and 8 RTX4090 GPUs. Our code is publicly available at https://github.com/bigdata-ustc/DMC-CDM.

4.3 Experimental Results

RQ1: Learners Performance Prediction To validate the effectiveness of DMC-CDM and the baselines, we first compare them in the Learners Performance Prediction task. Seen the results illustrated in Table 2, we can conclude that: (1) DMC-CDM significantly outperforms nearly all baseline methods across each dataset. This advancement can be attributed to the theoretical framework, which advocates for the integration of multi-perspective observations (Tewari et al. 2024). (2) NeuralCDM+ (Wang et al. 2023) and other variants generally outperform traditional IRT, MIRT, and NeuralCDM (Lord 1980; Ackerman, Gierl, and Walker 2003), demonstrating that incorporating expressive interaction function significantly enhances the ability to capture learners' cognitive states. This advantage persists even though these methods all overlook the issue of partial observations. (3) EIRS, CMES (Yao et al. 2023; Ma et al. 2024) demonstrates superiority over traditional CDMs through the incorporation of unobserved interactions through knowledge-aware sampling and the addition of pseudo labels. However, it is still worse than the proposed DMC-CDM, since it does not consolidate multi-perspective information in a theoretically sound way.

The aforementioned experimental phenomena provide partial evidence that the DMC-CDM and its constituent elements are efficacious.

RQ2: Ablation Study In order to validate the effectiveness of each component of the DMC-CDM, as summarized in RQ1, an ablation study was conducted. The DMC-CDM-w/o M employs solely the Single-perspective Extractor, whereas the DMC-CDM-w/o S consolidates potential collaborative representations. DMC-CDM-w/o Both utilizes collaborative representations derived from single-perspective extraction. The experimental results, presented in Table 2, illustrate the performance of our default method and its three variants across three datasets. From the results, we can conclude that: (1) Every variant performs better than DMC-CDM-w/o Both, indicating that components are effective in producing more precisely learners' cognitive states. (2) Multi-perspective consolidation plays a more crucial role than the expressive interaction function equipped with a semantic extractor. This is substantiated by the significantly superior performance of DMC-CDM-w/o S compared to DMC-CDM-w/o M. The reason may be that multi-perspective observations can provide sufficient information beyond the semantic information, compensating for information loss in the under-expressive interaction functions.

RQ3: Further Analysis In our theoretical analysis, we posit that observing all potential observations allows for the restoration of cognitive states with minimal information loss. In other words, more observations lead to better outcomes. However, in practice, we can only obtain potential observations from learners similar to the one undergoing assessment. As we retrieve more potential observations with a larger K, the less similar the tail-end learners are, resulting in diminishing returns from their interaction logs, and thus, less benefit from consolidating this information. Therefore, it is crucial to determine the optimal setting of K to achieve the best multi-perspective consolidation. We conducted an experimental analysis of the K setting on the Ifly dataset, and the results are presented in Figure 3. Our findings indicate that as K increases, both AUC and ACC gradually rise until reaching an upper bound. This confirms our hypothesis that beyond a certain threshold, interactions from more

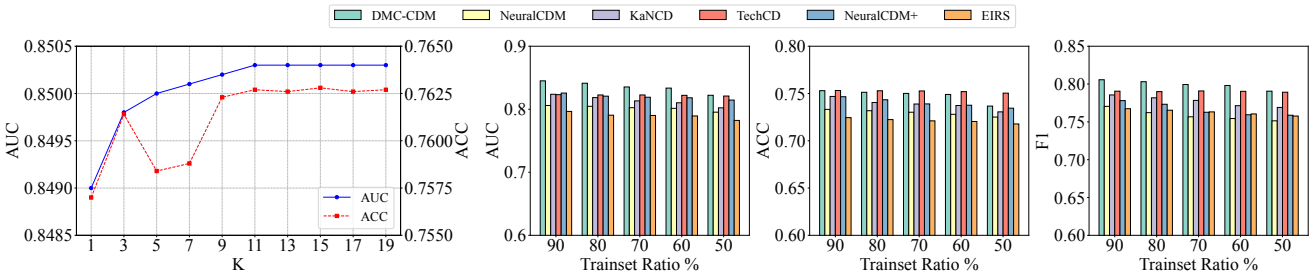

Figure 3: Analysis on the number of potential observations K and the robustness.

	MoocRadar		Junyi		Ifly	
	AUC ↑	ACC ↑	AUC ↑	ACC ↑	AUC ↑	ACC ↑
BGE-M3	0.9064	0.9143	0.8834	0.8597	0.8500	0.7614
Llama3-8B	0.8981	0.9157	0.8716	0.8545	0.8259	0.7450
Qwen2-7B	0.9030	0.9154	0.8695	0.8597	0.8363	0.7500
MiniCPM-2B	0.9010	0.9139	0.8814	0.8660	0.8393	0.7525
DMC-CDM-w/o S	0.9010	0.9116	0.8626	0.8537	0.8097	0.7301

Table 3: Impact of different semantic extractors.

dissimilar users do not contribute additional information to our multi-view fusion. The optimal K for best performance is determined to be 11, while a K of 3 provides a balance between performance and cost.

One step further, to assess the robustness of our proposed DMC-CDM and baselines in data-sparse scenarios, we conducted an experiments under varying degrees of sparsity. This was achieved by systematically removing random interactions from the *train set*. As illustrated in Figure 3, the performance of all CDMs exhibits a decline as the *train set* becomes progressively smaller relative to the original dataset, thereby increasing data sparsity. Notably, DMC-CDM consistently demonstrates superior performance compared to other baseline models across the majority of conditions. This suggests that the architecture of DMC-CDM confers enhanced robustness in sparse data environments.

Our proposed DMC-CDM model utilizes the semantic extractor to make sense of the semantic information, thus alleviating the information loss of under-expressive interaction functions. Therefore, the semantic comprehension ability of these models may also have an impact on the effectiveness of DMC-CDM. We conducted an experiment to apply different common open-source models to serve as the semantic extractor, including BGE-M3 (Chen et al. 2024a), Llama3-8B [4], Qwen2-7B (Yang et al. 2024a) and MiniCPM-2B (Hu et al. 2024). The results of the experiment are shown in Table 3, from which we can conclude that: (1) All variants performed better than the model without the semantic extractor, providing further evidence that using semantic information enhances learner modeling from a single perspective.

RQ4: Application Analysis Offering personalized learning recommendations based on the estimated cognitive states of learners is crucial for sustaining their motivation and enthusiasm, facilitating continuous engagement in learning, and ultimately enhancing their various abilities (Cheung et al. 2003). We conducted an experiment using a straightforward strategy (Lord 1980; Ghosh and Lan 2021), which recommends exercises that learners are expected to answer correctly with a probability of 0.5 (i.e., maximum information entropy, which is optimal for enhancing learners). This represents a moderate level of difficulty—not too hard, but not too easy—and provides the learner with the most benefit.

[4]https://github.com/meta-llama/llama3

Position		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(a)	Id	3578	2656	2688	2494	183	7977	3199	316	8266	1532
	Prob (%)	99.99	98.91	98.70	0.02	35.40	98.38	0.06	99.08	94.42	4.19
(b)	Id	3571	5634	6189	4716	7569	5878	190	4452	7160	9051
	Prob (%)	49.99	50.00	49.98	50.07	49.81	50.23	49.76	49.74	49.69	50.31

Table 4: The results on educational recommendation include: (a) the recommendation result using a random strategy, and (b) the recommendation result using our proposed strategy. "Id" is the id of the exercise, and "Prob(%)" indicates the probability of correctly answering.

To set up, we first randomly selected a learner and captured its cognitive states. Then, we recommended 10 exercises to the learner based on the aforementioned strategy. For comparison, we also randomly recommended 10 exercises to the learner. From Table 4, we can conclude that our strategy provides learners with exercises that are neither too difficult nor too easy. It offers the most informative and beneficial learning experience (Lord 2012; Ghosh and Lan 2021) rather than providing either too easy (e.g., 3578) or too difficult (e.g., 3199), which can hinder learner engagement and improvement (Huang et al. 2020).

5 Conclusion

In this paper, our focus was on alleviating the information loss, which made the cognitive diagnosis ill-posed and hindered the effectiveness of CDMs. We first established the theoretical basis for multi-perspective information consolidation. In line with this, we introduced a pioneering DMC-CDM. Specifically, we proposed a Single-perspective Extractor to leverage rich semantic information, enhancing the effectiveness of cognitive state assessments from a single perspective. Subsequently, we retrieved potential observations, maximized mutual information by conditional diffusion, thereby consolidating multi-perspective information. Our experimental results not only validated the efficacy of DMC-CDM but also explored its specific advantages.

Acknowledgments

This research was partially supported by grants from the Yangtze River Delta Science and Technology Innovation Community Joint Research Project (No.2023CSJZN0300), the National Natural Science Foundation of China (No.U20A20229, No.62477044, No.62106246), and the Key Technologies R & D Program of Anhui Province (No.202423k09020039).

References

Ackerman, T. A.; Gierl, M. J.; and Walker, C. M. 2003. Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement: Issues and Practice*, 22(3): 37–51.

Chen, J.; Xiao, S.; Zhang, P.; Luo, K.; Lian, D.; and Liu, Z. 2024a. Bge m3-embedding: Multilingual, multi-functionality, multi-granularity text embeddings through self-knowledge distillation. *arXiv preprint arXiv:2402.03216*.

Chen, X.; Wu, L.; Liu, F.; Chen, L.; Zhang, K.; Hong, R.; and Wang, M. 2024b. Disentangling cognitive diagnosis with limited exercise labels. *Advances in Neural Information Processing Systems*, 36.

Cheung, B.; Hui, L.; Zhang, J.; and Yiu, S.-M. 2003. SmartTutor: An intelligent tutoring system in web-based adult education. *Journal of Systems and Software*, 68(1): 11–25.

Chung, H.; Sim, B.; Ryu, D.; and Ye, J. C. 2022. Improving diffusion models for inverse problems using manifold constraints. *Advances in Neural Information Processing Systems*, 35: 25683–25696.

Dai, X.; Lin, J.; Zhang, W.; Li, S.; Liu, W.; Tang, R.; He, X.; Hao, J.; Wang, J.; and Yu, Y. 2021. An Adversarial Imitation Click Model for Information Retrieval. In *Proceedings of the Web Conference 2021*, 1809–1820.

De La Torre, J. 2009. DINA model and parameter estimation: A didactic. *Journal of educational and behavioral statistics*, 34(1): 115–130.

Gao, J.; Zhang, J.; Liu, X.; Darrell, T.; Shelhamer, E.; and Wang, D. 2023a. Back to the source: Diffusion-driven adaptation to test-time corruption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11786–11796.

Gao, W.; Liu, Q.; Huang, Z.; Yin, Y.; Bi, H.; Wang, M.-C.; Ma, J.; Wang, S.; and Su, Y. 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*, 501–510.

Gao, W.; Wang, H.; Liu, Q.; Wang, F.; Lin, X.; Yue, L.; Zhang, Z.; Lv, R.; and Wang, S. 2023b. Leveraging Transferable Knowledge Concept Graph Embedding for Cold-Start Cognitive Diagnosis. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 983–992.

Ghosh, A.; and Lan, A. 2021. BOBCAT: Bilevel Optimization-Based Computerized Adaptive Testing. In *International Joint Conference on Artificial Intelligence*.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33: 6840–6851.

Ho, J.; and Salimans, T. 2022. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*.

Hu, S.; Tu, Y.; Han, X.; He, C.; Cui, G.; Long, X.; Zheng, Z.; Fang, Y.; Huang, Y.; Zhao, W.; et al. 2024. Minicpm: Unveiling the potential of small language models with scalable training strategies. *arXiv preprint arXiv:2404.06395*.

Huang, Z.; Liu, Q.; Chen, Y.; Wu, L.; Xiao, K.; Chen, E.; Ma, H.; and Hu, G. 2020. Learning or Forgetting? A Dynamic Approach for Tracking the Knowledge Proficiency of Students. *ACM Trans. Inf. Syst.*, 38(2).

Jiang, L.; Wang, Y.; Wang, J.; Wang, P.; and Yin, M. 2023. Multi-View MOOC Quality Evaluation via Information-Aware Graph Representation Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(7): 8070–8077.

Jin, L.; Huang, W.; and Wen, Z. 2015. Developing a technoself system to improve lifelong learning engagement. In *2015 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, 102–107. IEEE.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Liu, J.; Huang, Z.; Xiao, T.; Sha, J.; Wu, J.; Liu, Q.; Wang, S.; and Chen, E. 2024. SocraticLM: Exploring Socratic Personalized Teaching with Large Language Models. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.

Lord, F. M. 1980. Applications of Item Response Theory To Practical Testing Problems.

Lord, F. M. 2012. *Applications of item response theory to practical testing problems*. Routledge.

Ma, H.; Wang, C.; Zhu, H.; Yang, S.; Zhang, X.; and Zhang, X. 2024. Enhancing cognitive diagnosis using un-interacted exercises: A collaboration-aware mixed sampling approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 8877–8885.

Nichol, A. Q.; and Dhariwal, P. 2021. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, 8162–8171. PMLR.

Peebles, W.; and Xie, S. 2023. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4195–4205.

Ren, X.; Wei, W.; Xia, L.; Su, L.; Cheng, S.; Wang, J.; Yin, D.; and Huang, C. 2024. Representation learning with large language models for recommendation. In *Proceedings of the ACM on Web Conference 2024*, 3464–3475.

Shen, L.; Zhao, W.; and Xing, L. 2019. Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning. *Nature Biomedical Engineering*, 880–888.

Song, Y.; Shen, L.; Xing, L.; and Ermon, S. 2022. Solving Inverse Problems in Medical Imaging with Score-Based

Generative Models. In *International Conference on Learning Representations*.

Song, Y.; Sohl-Dickstein, J.; Kingma, D. P.; Kumar, A.; Ermon, S.; and Poole, B. 2020. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*.

Tewari, A.; Yin, T.; Cazenavette, G.; Rezchikov, S.; Tenenbaum, J.; Durand, F.; Freeman, B.; and Sitzmann, V. 2024. Diffusion with forward models: Solving stochastic inverse problems without direct supervision. *Advances in Neural Information Processing Systems*, 36.

Wang, F.; Gao, W.; Liu, Q.; Li, J.; Zhao, G.; Zhang, Z.; Huang, Z.; Zhu, M.; Wang, S.; Tong, W.; and Chen, E. 2024. A Survey of Models for Cognitive Diagnosis: New Developments and Future Directions. arXiv:2407.05458.

Wang, F.; Liu, Q.; Chen, E.; Huang, Z.; Yin, Y.; Wang, S.; and Su, Y. 2023. NeuralCD: A General Framework for Cognitive Diagnosis. *IEEE Transactions on Knowledge and Data Engineering*, 35(8): 8312–8327.

Xiao, J.; Feng, R.; Zhang, H.; Liu, Z.; Yang, Z.; Zhu, Y.; Fu, X.; Zhu, K.; Liu, Y.; and Zha, Z.-J. 2024. DreamClean: Restoring Clean Image Using Deep Diffusion Prior. In *The Twelfth International Conference on Learning Representations*.

Yang, A.; Yang, B.; Hui, B.; Zheng, B.; Yu, B.; Zhou, C.; Li, C.; Li, C.; Liu, D.; Huang, F.; et al. 2024a. Qwen2 technical report. *arXiv preprint arXiv:2407.10671*.

Yang, Z.; Wu, J.; Wang, Z.; Wang, X.; Yuan, Y.; and He, X. 2024b. Generate what you prefer: Reshaping sequential recommendation via guided diffusion. *Advances in Neural Information Processing Systems*, 36.

Yao, F.; Liu, Q.; Hou, M.; Tong, S.; Huang, Z.; Chen, E.; Sha, J.; and Wang, S. 2023. Exploiting non-interactive exercises in cognitive diagnosis. *Interaction*, 100(200): 300.

Zhao, C.; Zhao, H.; He, M.; Zhang, J.; and Fan, J. 2023a. Cross-domain recommendation via user interest alignment. *Proceedings of the ACM Web Conference 2023*.

Zhao, G.; Huang, Z.; Zhuang, Y.; Bi, H.; Wang, Y.; Wang, F.; Ma, Z.; and Zhao, Y. 2024. A Diffusion-Based Cognitive Diagnosis Framework for Robust Learner Assessment. *IEEE Transactions on Learning Technologies*, 17: 2281–2295.

Zhao, G.; Huang, Z.; Zhuang, Y.; Liu, J.; Liu, Q.; Liu, Z.; Wu, J.; and Chen, E. 2023b. Simulating Student Interactions with Two-stage Imitation Learning for Intelligent Educational Systems. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, CIKM '23, 3423–3432. New York, NY, USA: Association for Computing Machinery. ISBN 9798400701245.