A Fast-Adaptive Cognitive Diagnosis Framework for Computerized Adaptive Testing Systems

Yuanhao Liu[1], **Yiya You**[1], **Shuo Liu**[1,2], **Hong Qian**[1*], **Ying Qian**[1] and **Aimin Zhou**[1]
[1]Shanghai Institute of AI Education, and School of Computer Science and Technology, East China Normal University, Shanghai 200062, China
[2]Game AI Center, Tencent Inc, Shenzhen 518057, China
{51275901044, 10225102493}@stu.ecnu.edu.cn, seokliu@tencent.com, {hqian, yqian, amzhou}@cs.ecnu.edu.cn

Abstract

Computerized Adaptive Testing (CAT) measures student ability by iteratively selecting informative questions, with core components being the Cognitive Diagnosis Model (CDM) and selection strategy. Current research focuses on optimizing the selection strategy, assuming relatively accurate CDM results. However, existing static CDMs struggle with rapid and accurate diagnosis in the early stage of CAT. To this end, this paper proposes a Fast Adaptive Cognitive Diagnosis (FACD) framework, which incorporates dynamic collaborative and personalized diagnosis modules. Specifically, the collaborative module in FACD uses a dynamic response graph to quickly build student cognitive profiles, while the personalized module leverages each student's response sequence for robust and individualized diagnosis. Extensive experiments on real-world datasets show that, compared with existing static CDMs, FACD not only achieves superior prediction performance across various selection strategies with an improvement between roughly 5%-10% in the early stage of CAT, but also maintains a commendable inference speed.

1 Introduction

With the rise of online testing platforms, the demand for personalized assessments has surged. Standardized exams like the GMAT and GRE have already used adaptive question selection strategies to better assess students' abilities. Computerized Adaptive Testing (CAT) [Mills and Steffen, 2000; Rudner, 2009; Liu *et al.*, 2024] optimizes test efficiency by tailoring questions to each student, reducing the number of questions needed while still ensuring accurate ability assessments. Unlike traditional tests, the Cognitive Diagnosis Model (CDM) in CAT estimates a student's current ability based on their responses to the questions selected by the selection strategy and provides feedback for further question selection. Consequently, CDM [Liu, 2021; Wang *et al.*, 2024; Li *et al.*, 2025] serves as a fundamental and decisive component in CAT [Anderson *et al.*, 2014].

*Corresponding Author.

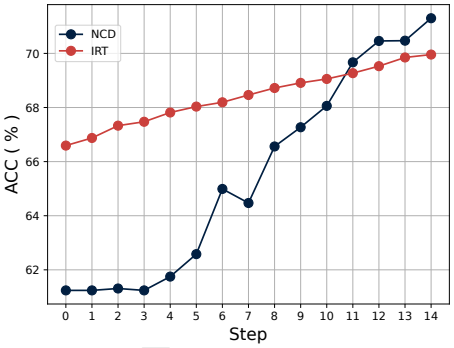


Figure 1: The motivation example. The accuracy of IRT and NCD on each step of CAT with BECAT selection strategy in EDM-Cup2023 dataset.

In recent decades, significant efforts have focused on developing CDMs, which model the interaction between students and questions. Traditional CDMs [Embretson and Reise, 2013; De La Torre, 2009; Sympson, 1978] rely on expert-crafted interaction functions, while neural network-based CDMs [Wang *et al.*, 2020a; Wang *et al.*, 2023a; Ma *et al.*, 2022; Li *et al.*, 2022; Liu *et al.*, 2023] use multilayer perceptrons to enhance generalization and interpretability. Recently, the use of graph neural networks (GNNs) to learn response relation graphs has gained attention, offering new insights for cognitive diagnosis [Qian *et al.*, 2024; Gao *et al.*, 2021; Shen *et al.*, 2024a; Wang *et al.*, 2023b].

Although the aforementioned CDMs have made remarkable progress, few studies focus on their adaptability to downstream tasks like CAT, which heavily relies on the assumption of relatively accurate CDM's outcomes. The goal of CAT is to assess a student's abilities with the fewest possible questions, posing a key challenge that ***CDM should be able to dynamically model students' abilities for rapid adaptation during the early stage of CAT, however the early stage suffers from limited data, which is not conducive to this rapid and accurate adaptation.*** As shown in Figure 1, we compare the performance of Item Response Theory (IRT) [Embretson and Reise, 2013] and Neural Cognitive Diagnosis Model (NCD) [Wang *et al.*, 2020a] in real CAT tasks using

BECAT [Zhuang *et al.*, 2023] for selection. IRT's simplicity limits its adaptability, resulting in slow performance improvements. Although NCD captures complex response patterns, its reliance on sufficient data leads to poor early-stage performance, conflicting with CAT's goal of rapid and accurate assessment.

To this end, this paper proposes a fast-adaptive cognitive diagnosis framework for computerized adaptive testing systems (FACD) to address the challenge mentioned above and effectively enhance CDM's adaptability especially during the early stage of CAT. Specifically, for the overall dynamic structure modeling, a dynamic collaborative diagnosis module are designed. Unlike traditional CDMs which rely on static response logs, the proposed FACD framework dynamically updates a response graph at each CAT stage to transfer relevant student and concept information, enabling rapid construction of the student's cognitive profile. For modeling personal dynamic response sequences, we design a dynamic personalized diagnosis module that employs a Gated Recurrent Unit (GRU) [Chung *et al.*, 2014] to capture the relationships in each student's response sequence. A self-attention mechanism further highlights critical response features. The collaborative information and the personalized information are then fused to model the ultimate representation and fed into the subsequent interaction function. Extensive experiments show that FACD improves early-stage performance in CAT by about 5%-10% compared with traditional CDMs.

The subsequent sections respectively recap the related work, present the preliminaries, introduce the proposed FACD, show the empirical analysis and conclude the paper. We include a section on the statement on the ethical use of data and informed consent of research subjects in Section 6.

2 Related Work

2.1 Computerized Adaptive Testing

CAT consists of two key components: a CDM and a selection strategy. Traditionally, IRT [Embretson and Reise, 2013] and NCD [Wang *et al.*, 2020a] are popular choices for CDM in CAT. Nowadays, most of the researches focus on the selection strategy in CAT. The most commonly used approaches like Maximum Fisher Information (MFI) [Lord, 2012] and Kullback-Leibler Information (KLI) [Chang and Ying, 1996] are typically tailored to specific CDMs, such as IRT. To address this limitation, Model-agnostic algorithms like MAAT [Bi *et al.*, 2020] and BECAT [Zhuang *et al.*, 2023] are introduced, but they rely on accurate CDM diagnosis, which is challenging in the early stages of CAT. Inaccurate CDM results can mislead selection strategies, delaying the whole CAT process. Data-driven strategies like BOBCAT [Ghosh and Lan, 2021] and NCAT [Zhuang *et al.*, 2022a] pretrain strategy models on large datasets, but still struggle with the instability of early-stage CDM diagnosis. Thus, a CDM tailored to CAT is essential for accurate and timely early-stage diagnosis.

2.2 Cognitive Diagnosis

Traditional cognitive diagnosis infers students' mastery levels of concepts using models like IRT [Embretson and Reise, 2013], MIRT [Sympson, 1978], and DINA [De La Torre, 2009]. With deep learning advances, models using multilayer perceptrons (MLPs) like NCD [Wang *et al.*, 2020a], KSCD [Ma *et al.*, 2022] and KaNCD [Wang *et al.*, 2023a] have successfully captured complex interactions. Other approaches to extracting rich information from response logs include symbolic regression [Shen *et al.*, 2024b], graph attention networks [Gao *et al.*, 2021; Qian *et al.*, 2024; Shen *et al.*, 2024a; Wang *et al.*, 2023b] and Bayesian networks [Li *et al.*, 2022]. Recently, several CDMs have been developed to address cold-start issues in static CD, such as TechCD [Gao *et al.*, 2023] and ZeroCD [Gao *et al.*, 2024], which focus on cross-domain CD, and BetaCD [Bi *et al.*, 2023] and AGCDM [Pei *et al.*, 2022], which tackle limited interaction data for specific students in CD. While these issues share some similarities with the cold-start problem in the early stage of CAT, they are not directly applicable due to dynamic context in CAT. Furthermore, few CDM studies focus on adapting to downstream tasks, and to our knowledge, no CDM has been specifically designed for CAT's unique requirements, making such adaptation crucial.

3 Preliminaries

Cognitive Diagnosis Basis. Consider an educational system with N students, M questions and K knowledge concepts, which can be denoted as three sets: $S = \{s_1, ..., s_N\}$, $E = \{e_1, ..., e_M\}$ and $C = \{c_1, ..., c_K\}$. The relationship between the questions and the knowledge concepts can be recorded in a binary Q-matrix $\mathbf{Q} = \{Q_{ij}\}_{M \times K}$ where $Q_{ij} \in \{0, 1\}$ indicates whether the knowledge concept is involved in the question or not. Each student in S will selectively do some questions from the set E, generating response logs that record their performance. These response logs are illustrated as triplets $R = \{(s, e, r) | s \in S, e \in E, r_{se} \in \{0, 1\}\}$ with $r_{se} = 1$ meaning question e is correctly answered by student s while $r_{se} = 0$ standing for the opposite. Based on the messages above, cognitive diagnosis aims to infer $Mas \in \mathbb{R}^{N \times K}$, a matrix that indicates each student's mastery level across all knowledge concepts.

Computerized Adaptive Testing. In CAT, the selection strategy and CDM alternate in a cyclical process. For the CDM, let us define the CDM as $\mathcal{M}$. And a student i's question set E_i can be divided as candidate and evaluated questions set as C_i and J_i. Assuming at time step $t \in [1, T]$, student i has already answered a sequence of questions $R_{t-1,i} = \{(e_1, r_1), ..., (e_{|t-1|}, r_{|t-1|})\}$ selected from candidate set C_i. Then, the CDM $\mathcal{M}$ can estimates student i's mastery level at time t based on the response sequence $R_{t-1,i}$, denoted as $\hat{Mas}_{t-1,i} = \mathcal{M}(R_{t-1,i}) \in [0, 1]$. For the selection strategy, let us define the selection policy as π which selects a question $e_{t,i} \sim \pi(\hat{Mas}_{t-1,i})$ for student i. Upon receiving student i's response $r_{t,i}$, the CDM is updated, resulting in a new mastery level estimate $\hat{Mas}_{t,i}$. And after updating, the performance of the CDM is tested on the evaluated questions set J_i. The above process is repeated T steps. The goal of CAT is to let the student estimated mastery level $\hat{Mas}_{T,i}$ close to the real value Mas_i^* with as few test steps T as possible.

Problem Statement. Our paper primarily focuses on building a fast-adaptive CDM module to help the CDM in CAT systems estimate a student's mastery level rapidly and stably, bringing it closer to the real value as soon as possible. Specifically, our goal is to ensure that at the relatively early time step T, the estimated mastery level $\hat{Mas}_{T,i} = \mathcal{M}(R_{T-1,i})$ is as close as possible to the real mastery level Mas_i^*. However, since we cannot directly observe Mas_i^*, our goal shifts to estimate the distribution of the student's actual responses R_i^*. The closer the estimated response distribution $R_{T,i}$ is to the real response distribution R_i^*, the better the CDM's performance. This difference is typically measured using Area Under the Curve (AUC), where a higher $\text{AUC}(\hat{R}_{T,i}, R_i^*)$ indicates more accurate predictions from the CDM.

4 Methodology

In this section, we provide the technical details of the proposed FACD, depicted in Figure 2. We introduce two crucial components: a dynamic collaborative diagnosis module in Figure 2(a) and a dynamic personalized diagnosis module in Figure 2(b). Following above, we will describe how to integrate these features into the interaction functions of various CDMs in the ultimate inference module. Detailed components of FACD are shown in Figure 2(c).

4.1 Dynamic Collaborative Diagnosis Module

Graph Construction. Firstly, we define the dynamic response graph at time step t for dynamic collaborative diagnosis module, denoted as $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}_t)$. The node set $\mathcal{V} = S \cup E \cup C$ involves students, questions, and concepts nodes respectively. $\mathcal{V}$ remains unchanged during the CAT process, with all nodes being defined initially. The edge set $\mathcal{E}_t = \mathcal{E}_t^{se} \cup \mathcal{E}^{ec}$ consists of two types of edges where $\mathcal{E}_t^{se}$ involves the interactions between students set S and questions set E (i.e., student s_i complete the question e_j) as well as $\mathcal{E}^{ec}$ involves the relationships between questions set E and concepts set C (i.e., question e_j is related to the concept c_k). Based on this, we can decompose the graph into two subgraphs. The first subgraph is the student-question graph, which is dynamic in nature. At each time step t, the edge set is updated as $\mathcal{E}_t^{se} = \mathcal{E}_{t-1}^{se} \cup R_t$, where R_t represents the newly added interaction records of all students at time step t. The corresponding adjacency matrix $\mathbf{A}_t^{se} \in \mathbb{R}^{|S|\times|E|}$, can be expressed as follows:

$$\mathbf{A}_t^{se} = \{x_{ij}\}_{|S|\times|E|}, \quad x_{ij} = \begin{cases} 1 & \text{if } (s_i, e_j, r_{ij}) \in \mathcal{E}_t^{se}, \\ 0 & \text{otherwise}. \end{cases} \tag{1}$$

The second subgraph is the question-concept graph, which is a static graph. The edge set $\mathcal{E}^{ec}$ can be directly derived from the relationships between questions and concepts in the Q-matrix. The corresponding matrix $\mathbf{A}^{ec} \in \mathbb{R}^{|E|\times|C|}$, can be expressed as $\mathbf{A}^{ec} = \mathbf{Q}$. Then we can construct the final dynamic response graph $\mathcal{G}_t$ which integrates the above two subgraphs. And its corresponding matrix $\mathbf{A}_t \in \mathbb{R}^{(|S|+|E|+|C|)\times(|S|+|E|+|C|)}$ can be expressed as follow:

$$\mathbf{A}_t = \begin{pmatrix} \mathbf{O} & \mathbf{A}_t^{se} & \mathbf{O} \\ \mathbf{A}_t^{se^\top} & \mathbf{O} & \mathbf{A}^{ec} \\ \mathbf{O} & \mathbf{A}^{ec^\top} & \mathbf{O} \end{pmatrix}, \tag{2}$$

where $\mathbf{O}$ means zero matrix. The graph $\mathcal{G}_t$ dynamically creates connections between students and questions during question selection in CAT and construct the relationship between question and concept, which is fundamental to the subsequent extraction of collaborative information.

Graph Convolution. In CDM, the fundamental data elements comprise response logs and the Q-matrix. It is necessary to systematically deconstruct these intricate logs into their essential components: students, questions, and concepts. To represent these entities, we employ trainable embeddings, denoted as $\mathbf{Z}_s^{(0)} \in \mathbb{R}^{|S|\times d}, \mathbf{Z}_e^{(0)} \in \mathbb{R}^{|E|\times d}, \mathbf{Z}_c^{(0)} \in \mathbb{R}^{|C|\times d}$ where d indicates the dimension of the embeddings, as initial ID representation. And to facilitate subsequent convolution processes, we stack the aforementioned embeddings to form $\mathbf{Z}^{(0)} \in \mathbb{R}^{(|S|+|E|+|C|)\times d}$, so the initial dynamic collaborative representation at time step $t = 0$ can be defined as $\mathbf{Z}_0^{(1)} = \mathbf{Z}^{(0)}$. Given that the representation of students, questions, and knowledge concepts in CDM are relatively straightforward and consist solely of IDs. Inspired from [He *et al.*, 2020], we eliminate linear transformations and nonlinear activation functions, choosing instead to focus on the fundamental elements of graph convolutional networks (GCNs). Therefore, the graph embedding propagation layer at time step t is formulated in the following matrix representation:

$$\mathbf{Z}_t^{(1,l)} = \left(\mathbf{D}^{-\frac{1}{2}}\mathbf{A_t}\mathbf{D}^{-\frac{1}{2}}\right)\mathbf{Z}_t^{(1,l-1)}, \tag{3}$$

where the degree matrix $\mathbf{D}$ is a diagonal matrix with size $(|S|+|E|+|C|)\times(|S|+|E|+|C|)$, and each entry $\mathbf{D}_{ii}$ indicates the number of non-zero entries in the i-th row vector of the adjacency matrix $\mathbf{A}_t$. By applying Eq. (3), we can derive the convolution results $\mathbf{Z}_t^{(1,l)} \in \mathbb{R}^{(|S|+|E|+|C|)\times d}$ from the l-th layer of graph $\mathcal{G}_t$ at time step t. The dynamic collaborative representation $\tilde{\mathbf{Z}}_t^{(1)} \in \mathbb{R}^{(|S|+|E|+|C|)\times d}$ at time step t is calculated through a mean pooling operation on the outcomes from each layer of graph $\mathcal{G}_t$ which can be expressed as $\tilde{\mathbf{Z}}_t^{(1)} = \frac{1}{1+L}\sum_{l=0}^{L}\mathbf{Z}_t^{(1,l)}$.

The graph convolution process aggregates outcomes from neighbors which is crucial for extracting collaborative information, as it enables us to identify patterns and relationships that reflect the collective strengths and weaknesses on student's mastery level within their similar neighbors. This helps us rapidly construct a profile for the student based on similar student groups in the early stage of CAT.

Graph Fusion. Given the inherent instability associated with training dynamic graphs, we fuse both the dynamic collaborative representation $\tilde{\mathbf{Z}}_t^{(1)} \in \mathbb{R}^{(|S|+|E|+|C|)\times d}$ of current time step t and the averaged dynamic collaborative representation $\overline{\mathbf{Z}}_t^{(1)} \in \mathbb{R}^{(|S|+|E|+|C|)\times d}$ from previous time steps 0 to $t-1$ in constructing the ultimate dynamic collaborative representation $\mathbf{Z}_t^{(1)} \in \mathbb{R}^{(|S|+|E|+|C|)\times d}$. This approach mitigates

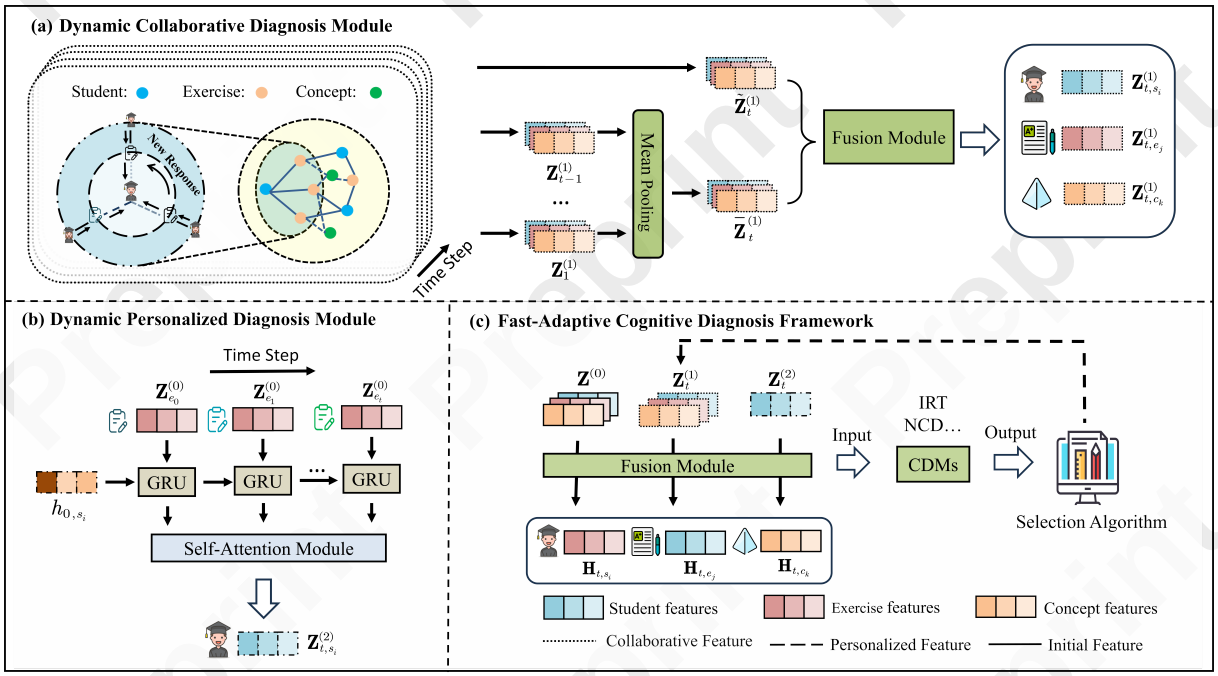


Figure 2: The overall framework of our proposed FACD. (a) Dynamic collaborative diagnosis module. (b) Dynamic personalized diagnosis module. (c) Detailed components of FACD.

the potential training instability arising from variations in the dynamic graph, which can be expressed as follows:

$$\mathbf{Z}_t^{(1)} = \eta_t \tilde{\mathbf{Z}}_t^{(1)} + (1-\eta_t)\overline{\mathbf{Z}}_t^{(1)}, \quad \overline{\mathbf{Z}}_t^{(1)} = \frac{1}{1+t}\sum_{i=0}^{t-1}\mathbf{Z}_i^{(1)}, \quad (4)$$

where η_t is the smoothing weight for the dynamic collaborative features which can either be set using a predefined threshold to determine the fusion ratio, or be adjusted adaptively through a learning mechanism. We currently set $\eta_t = 0.5$.

4.2 Dynamic Personalized Diagnosis Module

For every student s_i, we construct a dynamic personalized representation at time step t based on their response records of the selected questions. To capture the sequential relationships between questions chosen by the selection algorithm in CAT for each student, we utilize the GRU network [Chung *et al.*, 2014]. The personalized question representation sequences $\mathbf{L}_{t,s_i} \in \mathbb{R}^{t\times d}$ for student i at time step t are defined as follows:

$$\mathbf{L}_{t,s_i} = \{\mathbf{h}_{j,s_i}\}_{j=1}^{t}, \quad \mathbf{h}_{t,s_i} = \mathrm{GRU}(\mathbf{Z}_{e_t}^{(0)}, \mathbf{h}_{t-1,s_i}), \quad (5)$$

where $\mathbf{h}_{t,s_i} \in \mathbb{R}^{1\times d}$ denotes the hidden states of questions selected at the time step t in GRU. Then we utilize the multi-head self-attention over the personalized question representation sequences $\mathbf{L}_{t,s_i}$, to extract the most critical response information of each student and get the dynamic personalized representation $\mathbf{Z}_{t,s_i}^{(2)} \in \mathbb{R}^{1\times d}$ of student s_i at time step t:

$$\begin{aligned}\mathbf{Z}_{t,s_i,h}^{(2)} &= softmax\left(\frac{\mathbf{L}_{t,s_i}\mathbf{W}_h^{\mathbf{Q}}(\mathbf{L}_{t,s_i}\mathbf{W}_h^{\mathbf{K}})^{\mathsf{T}}}{\sqrt{d_h}}\right)\mathbf{L}_{t,s_i}\mathbf{W}_h^{\mathbf{V}}, \\ \mathbf{Z}_{t,s_i}^{(2)} &= [\mathbf{Z}_{t,s_i,1}^{(2)}, \mathbf{Z}_{t,s_i,2}^{(2)}, \ldots, \mathbf{Z}_{t,s_i,H}^{(2)}]\mathbf{W}^{\mathbf{F}} + b^{\mathbf{F}},\end{aligned} \quad (6)$$

where $\mathbf{W}_h^{\mathbf{Q}}, \mathbf{W}_h^{\mathbf{K}}, \mathbf{W}_h^{\mathbf{V}}, \in \mathbb{R}^{d\times d_h}$ are the trainable parameter matrices for the query matrix, key matrix and value matrix respectively, $\mathbf{W}^{\mathbf{F}} \in \mathbb{R}^{d\times d}, b^{\mathbf{F}} \in \mathbf{R}$ are the trainable parameters for the final linear transformation, $d_h = \frac{H}{d}$ denotes the dimension of each self-attention head, H stands for the number of self-attention heads, and $softmax(\cdot)$ and $[\cdot]$ is the row-wise softmax function and concatenation operation respectively.

Compared to traditional CDMs that treat the student, question and response records as discrete data points for training, our dynamic personalized diagnosis module offers a sequential modeling of questions for each student. This approach emphasizes the interrelationships among questions within the CAT selection strategy while extracting pertinent response information for each student using self-attention. Consequently, it lays a foundation for personalized modeling and enhances the robustness of CDMs in fast adaptation of CAT.

4.3 Inference Module

Representation Fusion. After obtaining the representations from the dynamic collaborative diagnosis module and dynamic personalized diagnosis module, the fusion module uniformly models the ultimate fused representation for each student s_i, each question e_j and each concept c_k at time step t. Specifically, we can express this fusion process as follow:

$$\begin{cases}\mathbf{H}_{t,s_i} = \alpha_{t,s_i}^0\mathbf{Z}_{s_i}^{(0)} + \alpha_{t,s_i}^1\mathbf{Z}_{t,s_i}^{(1)} + \alpha_{t,s_i}^2\mathbf{Z}_{t,s_i}^{(2)}, \\ \mathbf{H}_{t,e_j} = \alpha_{t,e_j}^0\mathbf{Z}_{e_j}^{(0)} + \alpha_{t,e_j}^1\mathbf{Z}_{t,e_j}^{(1)}, \\ \mathbf{H}_{t,c_k} = \alpha_{t,c_k}^0\mathbf{Z}_{c_k}^{(0)} + \alpha_{t,c_k}^1\mathbf{Z}_{t,c_k}^{(1)},\end{cases} \quad (7)$$

where $\mathbf{H}_{t,s_i} \in \mathbb{R}^{1\times d}$, $\mathbf{H}_{t,e_j} \in \mathbb{R}^{1\times d}$ and $\mathbf{H}_{t,c_k} \in \mathbb{R}^{1\times d}$ denote the ultimate fused representation for each student s_i,

Datasets	FrcSub	EDMCup2023	NeurIPS2020
#Students	536	27,178	4,918
#Questions	20	1,687	455
#Concepts	8	542	38
#Response Logs	10,720	413,021	494,316
Sparsity	1.000	0.009	0.221
Average Correct Rate	0.533	0.575	0.546
Q Density	2.800	1.000	1.000

Table 1: Statistics of the three real-world datasets.

each question e_j and each c_k at time step t respectively. $\mathbf{Z}_{s_i}^{(0)}$, $\mathbf{Z}_{t,s_i}^{(1)}$ and $\mathbf{Z}_{t,s_i}^{(2)}$ represent the initial ID representation, dynamic collaborative representation and dynamic personalized representation as mentioned above. And α_t is the normalized weight of three kinds of representation. Taking α_{t,s_i} as example, we first calculate these weights as:

$$\begin{cases} \alpha_{t,s_i}^0 = [\mathbf{Z}_{t,s_i}^{(0)}, \mathbf{Z}_{t,s_i}^{(1)}, \mathbf{Z}_{t,s_i}^{(2)}]\mathbf{W}_s^0 + b_s^0 , \\ \alpha_{t,s_i}^1 = [\mathbf{Z}_{t,s_i}^{(0)}, \mathbf{Z}_{t,s_i}^{(1)}, \mathbf{Z}_{t,s_i}^{(2)}]\mathbf{W}_s^1 + b_s^1 , \\ \alpha_{t,s_i}^2 = [\mathbf{Z}_{t,s_i}^{(0)}, \mathbf{Z}_{t,s_i}^{(1)}, \mathbf{Z}_{t,s_i}^{(2)}]\mathbf{W}_s^2 + b_s^2 , \end{cases} \tag{8}$$

where $\mathbf{W}_s^0 \in \mathbb{R}^{3d\times 1}$, $\mathbf{W}_s^1 \in \mathbb{R}^{3d\times 1}$, $\mathbf{W}_s^2 \in \mathbb{R}^{3d\times 1}$, $b_s^0 \in \mathbf{R}$, $b_s^1 \in \mathbf{R}$, $b_s^2 \in \mathbf{R}$ are trainable parameters with concatenation $[\cdot]$ for three kinds of representation. After that, we normalize the weights as follow:

$$\alpha_{t,s_i}^l = \frac{\exp\left(\alpha_{t,s_i}^l\right)}{\sum_{j=0}^{2}\exp\left(\alpha_{t,s_i}^j\right)} . \tag{9}$$

As for question and concept fused representations, the process is similar to the student.

Interaction Function. Given input representation $\mathbf{H}_{t,s_i}$, $\mathbf{H}_{t,e_j}$, $\mathbf{H}_{t,c_k}$, our model is capable of adapting to various interaction functions present in existing CDMs to predict the performance of students' practicing questions, which can be formulated as $\hat{y}_{ij} = \mathcal{M}(\phi_s(\mathbf{H}_{t,s_i}), \phi_e(\mathbf{H}_{t,e_j}), \phi_c(\mathbf{H}_{t,c}))$, where $\mathcal{M}(\cdot)$ can be the interaction functions in any kinds of CDMs, and $\phi(\cdot)$ is the transformation layer which facilitate the integration of representation with the majority of interaction functions in existing CDMs and help transform the representation's dimensions to suit the feature space in specific type of CDM. For simplicity, we use the linear transformation in our framework.

Model Training. In the CD task, we primarily employ the binary cross entropy loss function, which measures the difference between the model's predicted outcomes and the actual response scores within a mini-batch. And in CAT process, the loss is calculated on each student's response set R_t of the newly selected question set at time step t. The loss function can be formulated as:

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{|R_t|}\sum_{(i,j,r_{ij})\in R_t}^{|R_t|}[r_{ij}\log\hat{y}_{ij} + (1-r_{ij})\log(1-\hat{y}_{ij})] . \tag{10}$$

5 Experiments

5.1 Experimental Settings

Dataset Description. The experiments are conducted on three real-world datasets, i.e., FrcSub [DeCarlo, 2011; Tatsuoka, 1984], EDMCup2023 [Ethan Prihar, 2023] and NeurIPS2020 [Wang *et al.*, 2020b]. The details about datasets source and data preprocessing are depicted in the Appendix D.1. The Appendix A-D can be found in https://github.com/BW297/FACD.

Evaluation Metrics. Evaluating the performance of CDM in CAT systems is inherently difficult due to the challenges in accurately assessing real students' mastery level Mas^*. To tackle this issue, a common approach is to predict students' response scores. We measure the model's ability to predict whether a student successfully solves an question compared to the actual outcomes in the test set, utilizing standard classification metrics such as accuracy (ACC) and area under the curve (AUC).

Implementation Details. For parameter initialization, we employ Xavier [Glorot and Bengio, 2010], and for optimization purposes, Adam [Kingma and Ba, 2015] is adopted. The batch size is set within the range $\{32, 64, 128, 256\}$. The learning rate is chosen from $\{1e^{-3}, 3e^{-3}, 5e^{-3}, 7e^{-3}, 1e^{-2}\}$. The dimensions of the MLPs for all methods are consistent, being 512 and 256. To ensure robustness and reproducibility, each experiment is repeated ten times. The source code for our implementation can be found in https://github.com/BW297/FACD. More details of the implementation can be found in Appendix D.3.

5.2 Performance Comparison

To showcase the effectiveness of FACD, we integrate it with both IRT [Embretson and Reise, 2013] and NCD [Wang *et al.*, 2020a], which are widely utilized in CAT researches [Bi *et al.*, 2020; Zhuang *et al.*, 2022a; Zhuang *et al.*, 2023], as discussed in the previous sections. The detailed introduction of baselines used in experiments below is provided in Appendix D.2.

Performance Comparison with Different CAT Selection Strategies. To demonstrate the strong compatibility of our method with various CAT selection strategies, we have chosen a set of both classical and cutting-edge strategies as follows: Random, MAAT [Bi *et al.*, 2020], BOBCAT [Ghosh and Lan, 2021], NCAT [Zhuang *et al.*, 2022b] and BECAT [Zhuang *et al.*, 2023].

As shown in Table 2, it is evident that various CDMs demonstrate strong adaptability when integrated with FACD across all CAT selection strategies. Notably, during the initial stages of question selection, such as after answering approximately five questions, our model exhibits fast adaptation, creating a significant performance gap compared to the original CDMs. In later stages of question selection, the performance improvement of our model becomes more incremental, as it approaches the student's true mastery level. This highlights the model's ability to converge within a limited number of steps, providing precise feedback to CAT selection strategies and supporting a more efficient reduction in the number of questions selected.

Performance Comparison with Different Extraction Modules in CDM. To compare different representation extraction methods, we use BECAT as selection strategy and select several baselines for comparison with FACD as follows: RCD [Gao *et al.*, 2021] and ORCDF [Qian *et al.*, 2024].

Dataset		FrcSub				EDMCup2023				NeurIPS2020			
Metric		AUC (%)/ACC (%)											
Strategy	Step	IRT		NCD		IRT		NCD		IRT		NCD	
		OL	FA	OL	FA	OL	FA	OL	FA	OL	FA	OL	FA
Random	5	67.98/62.86	**70.49*/64.89***	72.02/59.76	**86.73*/78.96***	70.79/67.10	**77.06*/72.08***	73.49/61.59	**76.33*/70.81***	65.45/61.15	**72.04*/65.94***	65.03/61.06	**70.16*/64.60***
	10	76.95/71.21	**82.43*/74.54***	78.05/66.07	**89.35*/81.79***	72.29/68.37	**79.14*/73.40***	75.64/67.84	**79.61*/73.71***	65.70/61.29	**73.87*/67.38***	67.88/62.96	**72.95*/66.82***
	15	84.13/77.67	**88.02*/80.35***	85.75/77.23	**90.08*/82.55***	73.22/69.10	**80.51*/74.62***	76.99/70.99	**80.90*/74.66***	65.89/61.45	**74.73*/68.07***	71.19/65.04	**74.00*/67.67***
MAAT	5	69.22/61.99	**73.77*/65.89***	73.70/52.68	**87.22*/78.79***	71.45/67.80	**77.35*/72.49***	73.90/61.49	**77.52*/71.17***	66.64/62.10	**72.92*/66.77***	65.42/61.24	**70.41*/65.04***
	10	78.19/71.82	**86.85*/78.74***	77.87/68.43	**89.77*/81.07***	73.04/69.96	**80.30*/74.45***	76.08/66.30	**80.43*/74.41***	66.94/62.29	**74.68*/68.20***	68.66/63.78	**73.83*/67.63***
	15	84.93/77.95	**87.92*/79.89***	86.20/78.66	**89.82*/81.49***	74.05/68.94	**81.61*/75.62***	77.26/70.72	**81.68*/75.69***	67.23/62.34	**75.48*/68.89***	71.58/65.74	**74.96*/68.65***
BOBCAT	5	73.84/67.21	**74.97*/65.46***	73.83/66.17	**87.36*/78.00***	71.58/67.68	**77.79*/72.16***	75.00/66.34	**77.15*/71.03***	66.77/62.11	**72.18*/66.40***	65.93/60.89	**70.76*/65.20***
	10	78.93/71.96	**87.75*/80.78***	78.67/68.83	**90.19*/80.87***	72.57/68.48	**79.81*/73.76***	76.81/69.19	**79.43*/73.37***	67.63/62.61	**74.43*/67.98***	68.52/63.67	**73.47*/67.29***
	15	82.11/74.60	**89.82*/82.37***	84.82/76.57	**90.93*/82.50***	73.70/69.37	**81.92*/75.94***	78.47/69.32	**81.24*/74.93***	68.36/62.98	**75.42*/68.83***	71.44/65.89	**74.64*/68.38***
NCAT	5	68.46/61.51	**70.87*/63.78***	73.21/52.68	**87.19*/79.04***	71.80/68.13	**77.70*/72.60***	74.40/62.59	**76.35*/70.95***	66.60/62.07	**72.80*/66.76***	65.95/60.82	**70.61*/65.24***
	10	78.55/72.42	**88.12*/80.64***	78.16/66.51	**90.17*/82.81***	73.48/69.25	**80.11*/74.37***	76.68/69.23	**79.83*/74.09***	66.95/62.23	**74.72*/68.30***	67.87/63.19	**73.75*/67.76***
	15	85.37/78.10	**90.50*/82.64***	84.74/76.00	**90.87*/83.37***	74.41/70.20	**81.51*/75.46***	78.07/71.58	**81.40*/75.62***	67.19/62.45	**75.50*/69.02***	71.06/65.37	**74.72*/68.65***
BECAT	5	70.87/63.29	**71.62*/65.07***	72.92/52.68	**85.97*/73.17***	71.82/68.03	**77.18*/72.67***	74.45/62.58	**77.48*/72.39***	66.40/62.17	**72.48*/66.52***	65.22/61.72	**70.64*/64.90***
	10	79.11/71.97	**88.03*/80.64***	77.98/67.33	**89.77*/81.61***	73.18/69.06	**81.03*/75.27***	76.67/68.06	**80.08*/74.45***	66.63/62.27	**74.56*/68.08***	67.96/63.13	**73.46*/67.39***
	15	84.85/78.21	**90.42*/82.58***	85.66/79.94	**90.49*/81.91***	74.12/69.96	**82.49*/76.45***	78.14/71.30	**81.44*/75.35***	66.77/62.28	**75.43*/68.86***	72.06/65.34	**74.63*/68.32***

Table 2: Overall performance with different CAT selection strategies. "OL" stands for "original", referring to the original method, and "FA" denotes the proposed FACD enhancement applied to the original method. Other details are as same as Table 3.

Dataset	FrcSub						EDMCup2023						NeurIPS2020					
IF	IRT			NCD			IRT			NCD			IRT			NCD		
Metric	AUC (%)																	
Step	5	10	15	5	10	15	5	10	15	5	10	15	5	10	15	5	10	15
OR	69.09	77.98	89.76	76.80	85.85	89.62	74.64	78.70	79.96	74.51	76.49	78.18	69.15	70.93	72.30	68.05	69.77	72.18
RCD	66.84	75.59	86.83	73.52	80.17	86.64	75.70	80.01	82.07	72.97	76.19	78.19	70.54	73.83	74.66	66.81	68.33	70.69
FA	**71.62***	**88.03***	**90.42**	**85.97***	**89.78***	**90.49***	**77.18**	**81.03**	**82.49**	**76.95***	**80.53***	**82.01***	**72.48**	**74.57**	**75.43**	**70.64***	**73.46***	**74.63***
Metric	ACC (%)																	
Step	5	10	15	5	10	15	5	10	15	5	10	15	5	10	15	5	10	15
OR	62.59	68.71	78.59	66.73	75.30	79.12	69.85	72.64	73.57	69.34	71.11	71.70	62.34	64.48	65.60	63.27	63.87	]65.76
RCD	60.94	71.07	82.02	63.90	70.43	75.69	70.60	73.62	75.75	68.44	70.78	71.77	64.91	67.49	68.10	62.26	63.31	64.60
FA	**65.07***	**80.14***	**82.58**	**73.17***	**81.61***	**81.91**	**72.67**	**75.27**	**76.45**	**70.62***	**74.35***	**75.41***	**66.52***	**68.08**	**68.86***	**64.90***	**67.39***	**68.32***

Table 3: Overall performance with extraction modules in IRT and NCD using BECAT. "FA" denotes the proposed FACD. "OR" denotes ORCDF. "IF" denotes interaction function. In each column, an entry with the best mean value is marked in bold and underline for the runner-up. The standard deviation is not shown in the table since it is very low (less than 0.05). If the mean value of the best model significantly differs from the runner-up, passing a t-test with a significance level of 0.05, then we denote it with "*" at the corresponding position.

As shown in Table 3, our FACD consistently outperforms RCD and ORCDF in CAT system and demonstrates relatively stable results across different CDMs. This is primarily because RCD and ORCDF lack stability research for dynamic graph modules. Furthermore, they fail to effectively model the personalized sequence of questions for each student and thus do not assess the importance of student response information, making them sensitive to noise, such as low-quality response data, which reduces their robustness. It can be observed that FACD not only ensures strong performance and robustness but also effectively mitigates BECAT's cold start issue as mentioned in original paper [Zhuang *et al.*, 2023]. We also provide further comparison with traditional classical GNN frameworks in Appendix D.5.

5.3 Ablation Study

To showcase the contributions of each component in FACD, we conduct an ablation study on FACD in dataset EDMCup2023, which is divided into the following three versions. "FA-C" and "FA-P" means removing dynamic collaborative diagnosis module and dynamic personalized diagnosis module, respectively. "OL" means the base CDMs. As shown in Table 4, the proposed method outperforms the other two ablated versions, indicating that each component contributes significantly to different stages of CAT, resulting in the overall effectiveness of the model. Early performance relies on the dynamic collaborative diagnosis, while later stability comes from the dynamic personalized diagnosis.

Dataset	EDMCup2023							
Strategy	Step	IRT				NCD			
		OL	FA-C	FA-P	FA	OL	FA-C	FA-P	FA
Random	5	70.79	75.62	75.43	**77.06**	73.50	74.76	75.75	**76.34**
	10	72.30	77.58	77.01	**79.14**	75.65	78.01	79.15	**79.62**
	15	73.23	78.78	79.31	**80.51**	77.00	80.63	80.68	**80.91**
MAAT	5	71.45	76.67	75.61	**77.36**	73.91	75.02	76.54	**77.53**
	10	73.04	79.50	77.63	**80.31**	76.08	78.75	79.74	**80.44**
	15	74.06	80.40	79.23	**81.61**	77.26	80.96	80.70	**81.69**
NCAT	5	71.80	76.47	75.85	**77.71**	74.41	75.14	75.55	**76.36**
	10	73.48	79.31	79.21	**80.11**	76.69	78.63	79.29	**79.83**
	15	74.42	80.91	79.84	**81.52**	78.08	80.92	80.67	**81.41**
BECAT	5	71.83	77.22	76.59	**77.48**	74.46	75.57	75.70	**76.95**
	10	73.18	79.70	79.01	**80.09**	76.68	78.42	79.36	**80.53**
	15	74.12	81.06	80.05	**81.45**	78.15	80.58	81.29	**82.01**

Table 4: Overall performance of ablation study in EDMCup2023 dataset. Details are as same as Table 3.

5.4 In-Depth Analysis of FACD's Advantages

Inference Time Comparison. We compare the inference speeds of different extraction modules and the original CDMs on the EDMCup2023 dataset, specifically the average CPU time per round spent on training and inference after the CAT

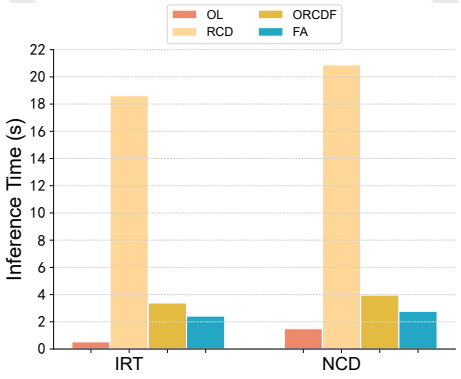
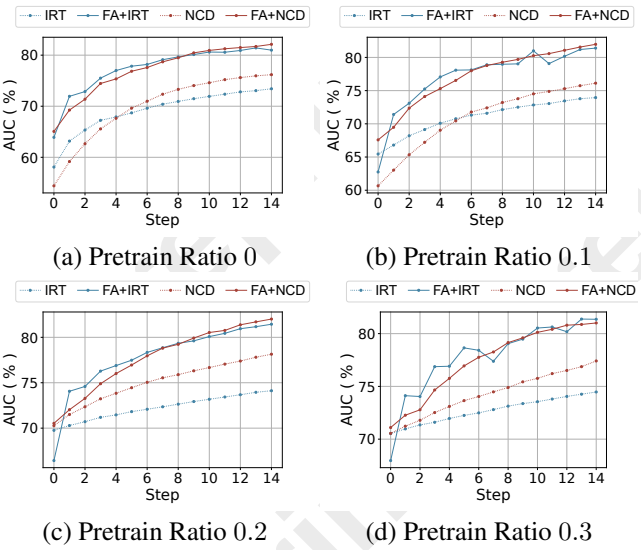


Figure 3: Inference time comparison.



(a) Pretrain Ratio 0 (b) Pretrain Ratio 0.1

(c) Pretrain Ratio 0.2 (d) Pretrain Ratio 0.3

Figure 4: Pretraining ratio analysis.

system selects a question for each student. As shown in Figure 3, RCD, using complex graph aggregation functions, has slower inference speeds and higher resource consumption. Both FACD and ORCDF, using lighter aggregation methods similar to LightGCN [He *et al.*, 2020], are faster, but ORCDF incurs additional costs due to its more complex response graph. Our method achieves high-quality cognitive diagnosis with limited computational resources, making it well-suited for modern online education platforms. Detailed time and space complexity analysis is provided in Appendix B.2.

Generalization Analysis on Pretrain Ratio. We evaluate the performance of the original CDMs on the EDM-Cup2023 dataset with FACD integration at different pretrain ratios $p_t = \{0.0, 0.1, 0.2, 0.3\}$, which represent the portion of the dataset used for pretraining before the CAT process. As shown in Figure 4, the original CDMs show performance drops with smaller pretrain ratios and require more steps in the CAT process. However, after integrating FACD, models quickly improve their performance in subsequent question selection steps, helping CDMs better align with students' real mastery levels.

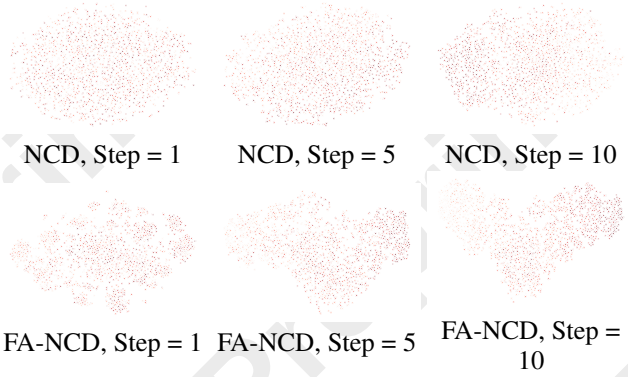


Figure 5: t-SNE scatter plots for the distribution shift of students' Mas on NCD and FA-NCD.

The Distribution Shift of Students' Mas. Students can naturally be grouped based on their performance, such as high or low correct response rates, reflecting differences in their mastery level. To visualize this, we use t-SNE [Van der Maaten and Hinton, 2008] to project Mas onto a two-dimensional plane (the 2D axes in t-SNE are for visualization only and hold no specific meaning), shading the scatter plot based on correct rates, where darker shades indicate higher correct rates. In Figure 5, the student representations learned by NCD with FACD on the NeurIPS2020 dataset exhibit clear clustering by the fifth step, becoming more distinct by the tenth step. In contrast, the standard NCD model shows no clear clustering at the fifth step, only displaying faint groupings by the tenth step. This demonstrates FACD's rapid adaptability in both representation learning and interpretability. Besides the visualization to show our model's interpretability, we also provide further comparison on the quantifiable interpretability metrics, such as DOA, in Appendix D.4.

5.5 Hyperparameter Analysis

We study the impact of the hyperparameters on the dynamic graph layer, GRU Layer, embedding dimension and graph mask ratio on the EDMcup2023 dataset. Specially, the model achieves optimal performance when the number of graph layers is set at 1 and 2, the number of GRU layers is set to 3, the embedding dimension is set to 32 or 64 and the graph mask ratio is set to 0.1 and 0.2. The detailed hyperparameter analysis and figure is provided in Appendix D.6.

6 Conclusion

In this paper, we propose a fast adaptive cognitive diagnosis framework called FACD for computerized adaptive testing. FACD focuses mainly on addressing the slow improvement issue of CDM in the early stage of CAT. Two dynamic diagnosis modules are introduced to capture collaborative information from similar students and critical information from individual response sequences, achieving fast adaptability. FACD still has some limitations, e.g., a lack of sufficient theoretical analysis to ensure effective alignment with CAT question selection strategies. Thus, future research could focus on strengthening the theoretical basis to enhance the interaction between CD and CAT.

Ethical Statement

The research presented does not involve human subjects or raise concerns related to privacy, security, or legal compliance. The datasets used in this study are from publicly available datasets that have undergone rigorous ethical review and anonymization during their release. These datasets contain no private student information and are widely used in various educational research studies.

Acknowledgments

The authors would like to thank the anonymous reviewers for their constructive and helpful comments. This work is supported by the National Natural Science Foundation of China (No. 62476091) and the National Undergraduate Training Program on Innovation and Entrepreneurship Grant (No. 202510269105G).

References

[Anderson *et al.*, 2014] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. Engaging with massive online courses. In *Proceedings of the 23rd international conference on World Wide Web*, pages 687–698, Seoul, Korea, 2014.

[Bi *et al.*, 2020] Haoyang Bi, Haiping Ma, Zhenya Huang, Yu Yin, Qi Liu, Enhong Chen, Yu Su, and Shijin Wang. Quality meets diversity: A model-agnostic framework for computerized adaptive testing. In *Proceedings of the 20th IEEE International Conference on Data Mining*, pages 42–51, Sorrento, Italy, 2020.

[Bi *et al.*, 2023] Haoyang Bi, Enhong Chen, Weidong He, Han Wu, Weihao Zhao, Shijin Wang, and Jinze Wu. Bettacd: A bayesian meta-learned cognitive diagnosis framework for personalized learning. In *Proceedings of the 37th AAAI conference on artificial intelligence*, volume 37, pages 110–118, Washington, DC, 2023.

[Chang and Ying, 1996] Hua-Hua Chang and Zhiliang Ying. A global information approach to computerized adaptive testing. *Applied Psychological Measurement*, 20(3):213–229, 1996.

[Chung *et al.*, 2014] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.

[De La Torre, 2009] Jimmy De La Torre. Dina model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, 34(1):115–130, 2009.

[DeCarlo, 2011] Lawrence T DeCarlo. On the analysis of fraction subtraction data: The dina model, classification, latent class sizes, and the q-matrix. *Applied Psychological Measurement*, 35(1):8–26, 2011.

[Embretson and Reise, 2013] Susan E Embretson and Steven P Reise. *Item Response Theory*. Psychology Press, 2013.

[Ethan Prihar, 2023] NHeffernan Ethan Prihar. Edm cup 2023, 2023.

[Gao *et al.*, 2021] Weibo Gao, Qi Liu, Zhenya Huang, Yu Yin, Haoyang Bi, Mu-Chun Wang, Jianhui Ma, Shijin Wang, and Yu Su. 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*, pages 501–510, Virtual Event, 2021.

[Gao *et al.*, 2023] Weibo Gao, Hao Wang, Qi Liu, Fei Wang, Xin Lin, Linan Yue, Zheng Zhang, Rui Lv, and Shijin Wang. 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*, pages 983–992, Taiwan, China, 2023.

[Gao *et al.*, 2024] Weibo Gao, Qi Liu, Hao Wang, Linan Yue, Haoyang Bi, Yin Gu, Fangzhou Yao, Zheng Zhang, Xin Li, and Yuanjing He. Zero-1-to-3: Domain-level zero-shot cognitive diagnosis via one batch of early-bird students towards three diagnostic objectives. In *Proceedings of the 38th AAAI Conference on Artificial Intelligence*, pages 8417–8426, Vancouver,Canada, 2024.

[Ghosh and Lan, 2021] Aritra Ghosh and Andrew Lan. Bobcat: Bilevel optimization-based computerized adaptive testing. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence*, Virtual Event, 2021.

[Glorot and Bengio, 2010] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*, pages 249–256, Sardinia, Italy, 2010.

[He *et al.*, 2020] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yong-Dong Zhang, and Meng Wang. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*, pages 639–648, Virtual Event, 2020.

[Kingma and Ba, 2015] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations*, San Diego, California, 2015.

[Li *et al.*, 2022] Jiatong Li, Fei Wang, Qi Liu, Mengxiao Zhu, Wei Huang, Zhenya Huang, Enhong Chen, Yu Su, and Shijin Wang. HierCDF: A Bayesian network-based hierarchical cognitive diagnosis framework. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 904–913, Virtual Event, 2022.

[Li *et al.*, 2025] Mingjia Li, Hong Qian, Jinglan Lv, Mengliang He, Wei Zhang, and Aimin Zhou. Foundation model enhanced derivative-free cognitive diagnosis. *Frontiers Comput. Sci.*, 19(1):191318, 2025.

[Liu *et al.*, 2023] Shuo Liu, Hong Qian, Mingjia Li, and Aimin Zhou. QCCDM: A q-augmented causal cognitive diagnosis model for student learning. In *Proceedings of*

the 26th European Conference on Artificial Intelligence, pages 1536–1543, Kraków, Poland, 2023.

[Liu *et al.*, 2024] Qi Liu, Yan Zhuang, Haoyang Bi, Zhenya Huang, Weizhe Huang, Jiatong Li, Junhao Yu, Zirui Liu, Zirui Hu, Yuting Hong, et al. Survey of computerized adaptive testing: A machine learning perspective. *arXiv preprint arXiv:2404.00712*, 2024.

[Liu, 2021] Qi Liu. Towards a new generation of cognitive diagnosis. In *Proceedings of 30th International Joint Conference on Artificial Intelligence*, pages 4961–4964, Montreal, Canada, 2021.

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

[Ma *et al.*, 2022] Haiping Ma, Manwei Li, Le Wu, Haifeng Zhang, Yunbo Cao, Xingyi Zhang, and Xuemin Zhao. Knowledge-sensed cognitive diagnosis for intelligent education platforms. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 1451–1460, Atlanta, GA, 2022.

[Mills and Steffen, 2000] Craig N Mills and Manfred Steffen. The gre computer adaptive test: Operational issues. In *Computerized adaptive testing: Theory and practice*, pages 75–99. Springer, 2000.

[Pei *et al.*, 2022] Xiaohuan Pei, Shuo Yang, Jiajun Huang, and Chang Xu. Self-attention gated cognitive diagnosis for faster adaptive educational assessments. In *Proceedings of the 22nd IEEE International Conference on Data Mining*, pages 408–417, Orlando, FL, 2022.

[Qian *et al.*, 2024] Hong Qian, Shuo Liu, Mingjia Li, Bingdong Li, Zhi Liu, and Aimin Zhou. Orcdf: An oversmoothing-resistant cognitive diagnosis framework for student learning in online education systems. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2455–2466, Barcelona, Spain, 2024.

[Rudner, 2009] Lawrence M Rudner. Implementing the graduate management admission test computerized adaptive test. In *Elements of adaptive testing*, pages 151–165. Springer, 2009.

[Shen *et al.*, 2024a] Junhao Shen, Hong Qian, Shuo Liu, Wei Zhang, Bo Jiang, and Aimin Zhou. Capturing homogeneous influence among students: Hypergraph cognitive diagnosis for intelligent education systems. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2628–2639, Barcelona, Spain, 2024.

[Shen *et al.*, 2024b] Junhao Shen, Hong Qian, Wei Zhang, and Aimin Zhou. Symbolic cognitive diagnosis via hybrid optimization for intelligent education systems. In *Proceedings of the 38th AAAI Conference on Artificial Intelligence*, pages 14928–14936, Vancouver, Canada, 2024.

[Sympson, 1978] James B Sympson. A model for testing with multidimensional items. In *Proceedings of the 1977 Computerized Adaptive Testing Conference*, Minneapolis, MN, 1978.

[Tatsuoka, 1984] Kikumi K Tatsuoka. Analysis of errors in fraction addition and subtraction problems. final report. 1984.

[Van der Maaten and Hinton, 2008] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(11), 2008.

[Wang *et al.*, 2020a] Fei Wang, Qi Liu, Enhong Chen, Zhenya Huang, Yuying Chen, Yu Yin, Zai Huang, and Shijin Wang. Neural cognitive diagnosis for intelligent education systems. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, New York, NY, 2020.

[Wang *et al.*, 2020b] Zichao Wang, Angus Lamb, Evgeny Saveliev, Pashmina Cameron, Yordan Zaykov, José Miguel Hernández-Lobato, Richard E Turner, Richard G Baraniuk, Craig Barton, Simon Peyton Jones, et al. Instructions and guide for diagnostic questions: The neurips 2020 education challenge. *arXiv preprint arXiv:2007.12061*, 2020.

[Wang *et al.*, 2023a] Fei Wang, Qi Liu, Enhong Chen, Zhenya Huang, Yu Yin, Shijin Wang, and Yu Su. Neuralcd: A general framework for cognitive diagnosis. *IEEE Transactions on Knowledge and Data Engineering*, 35(8), 2023.

[Wang *et al.*, 2023b] Shanshan Wang, Zhen Zeng, Xun Yang, and Xingyi Zhang. Self-supervised graph learning for long-tailed cognitive diagnosis. In *Proceedings of the 37th AAAI conference on artificial intelligence*, volume 37, pages 110–118, Washington, DC, 2023.

[Wang *et al.*, 2024] Fei Wang, Weibo Gao, Qi Liu, Jiatong Li, Guanhao Zhao, Zheng Zhang, Zhenya Huang, Mengxiao Zhu, Shijin Wang, Wei Tong, et al. A survey of models for cognitive diagnosis: New developments and future directions. *arXiv preprint arXiv:2407.05458*, 2024.

[Zhuang *et al.*, 2022a] Yan Zhuang, Qi Liu, Zhenya Huang, Zhi Li, Shuanghong Shen, and Haiping Ma. Fully adaptive framework: Neural computerized adaptive testing for online education. In *Proceeddings of the 36th AAAI Conference on Artificial Intelligence*, pages 4734–4742, Virtual Event, 2022.

[Zhuang *et al.*, 2022b] Yan Zhuang, Qi Liu, Zhenya Huang, Zhi Li, Shuanghong Shen, and Haiping Ma. Fully adaptive framework: Neural computerized adaptive testing for online education. In *Proceedings of the 36th AAAI conference on artificial intelligence*, volume 36, pages 4734–4742, 2022.

[Zhuang *et al.*, 2023] Yan Zhuang, Qi Liu, GuanHao Zhao, Zhenya Huang, Weizhe Huang, Zachary Pardos, Enhong Chen, Jinze Wu, and Xin Li. A bounded ability estimation for computerized adaptive testing. In *Advances in Neural Information Processing Systems 36*, New Orleans, LA, 2023.