

# Fuzzy Cognitive Diagnosis for Modelling Examinee Performance

QI LIU, RUNZE WU, and ENHONG CHEN, University of Science and Technology of China  
GUANDONG XU, University of Technology, Sydney  
YU SU, Anhui University & IFLYTEK Co., Ltd.  
ZHIGANG CHEN and GUOPING HU, IFLYTEK Co., Ltd.

Recent decades have witnessed the rapid growth of educational data mining (EDM), which aims at automatically extracting valuable information from large repositories of data generated by or related to people's learning activities in educational settings. One of the key EDM tasks is cognitive modelling with examination data, and cognitive modelling tries to profile examinees by discovering their latent knowledge state and cognitive level (e.g. the proficiency of specific skills). However, to the best of our knowledge, the problem of extracting information from both objective and subjective examination problems to achieve more precise and interpretable cognitive analysis remains underexplored. To this end, we propose a *fuzzy cognitive diagnosis framework* (FuzzyCDF) for examinees' cognitive modelling with both objective and subjective problems. Specifically, to handle the partially correct responses on subjective problems, we first fuzzify the skill proficiency of examinees. Then we combine fuzzy set theory and educational hypotheses to model the examinees' mastery on the problems based on their skill proficiency. Finally, we simulate the generation of examination score on each problem by considering slip and guess factors. In this way, the whole diagnosis framework is built. For further comprehensive verification, we apply our FuzzyCDF to three classical cognitive assessment tasks, i.e., predicting examinee performance, slip and guess detection, and cognitive diagnosis visualization. Extensive experiments on three real-world datasets for these assessment tasks prove that FuzzyCDF can reveal the knowledge states and cognitive level of the examinees effectively and interpretatively.

CCS Concepts: • **Computing methodologies** → **Learning in probabilistic graphical models**; *Vagueness and fuzzy logic*; • **Applied computing** → **Education**;

Additional Key Words and Phrases: Cognitive, graphic model, educational data mining

## ACM Reference format:

Qi Liu, Runze Wu, Enhong Chen, Guandong Xu, Yu Su, Zhigang Chen, and Guoping Hu. 2018. Fuzzy Cognitive Diagnosis for Modelling Examinee Performance. *ACM Trans. Intell. Syst. Technol.* 9, 4, Article 48 (January 2018), 26 pages.

<https://doi.org/10.1145/3168361>

This research was supported by grants from the National Natural Science Foundation of China (Grants No. 61672483, U1605251, 61325010, and 61727809) and Australian Research Council (ARC) Linkage Project under LP140100937. Qi Liu gratefully acknowledges the support of the Youth Innovation Promotion Association of CAS (No. 2014299).

Authors' addresses: Q. Liu, R. Wu, and E. Chen (corresponding author), Anhui Province Key Laboratory of Big Data Analysis and Application, University of Science and Technology of China; emails: qiliuql@ustc.edu.cn, cheneh@ustc.edu.cn, wrz179@mail.ustc.edu.cn; G. Xu, Advanced Analytics Institute and School of Software, University of Technology, Sydney; email: Guandong.Xu@uts.edu.au; Y. Su, School of Computer Science and Technology, Anhui University and iFLYTEK Co., Ltd.; email: yusu@iflytek.com; Z. Chen and G. Hu, iFLYTEK Co., Ltd.; emails: zgchen@iflytek.com, gphu@iflytek.com.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2018 ACM 2157-6904/2018/01-ART48 \$15.00

<https://doi.org/10.1145/3168361>

## 1 INTRODUCTION

During the past few decades, one of the most important innovations in computer aided education has been educational data mining (EDM), which is designed for exploiting valuable information from large-scale repositories of collected data with educational settings (Ma et al. 2000; Beck and Woolf 2000). Involving the applications of data mining, machine learning, psychology, and statistics, EDM has drawn a great deal of attention in various educational scenarios, e.g., massive open on-line courses (MOOCs) (Anderson et al. 2014), intelligent tutoring systems (Burns et al. 2014), and cognitively diagnostic assessment (CDA) (Nichols et al. 2012).

As the main participants in an education context, learners in different scenarios are usually targeted for modelling or profiling. For instance, the examinees, i.e., learners in CDA, are modelled based on examination data for discovering their knowledge state and cognitive level (e.g., the proficiency of specific skills), and this is referred to as *cognitive modelling*. With the comprehensive understanding of examinees, cognitive modelling could be further applied to numerous applications, such as personalized remedy recommendation and teaching plan improvement. Therefore, massive efforts in both psychometrics and data mining have been undertaken to improve the solutions of cognitive modelling: In psychometrics, most of the existing studies focus on cognitive diagnosis. In the cognitive diagnosis models (CDMs) (DiBello et al. 2006), examinees are characterized by the proficiency of specific skills (e.g., problem-solving skills like *calculation*), and typically a Q-matrix (Tatsuoka 1984) is given as the prior knowledge from education experts for denoting which skills are needed for each problem. In terms of data mining, matrix factorization (MF) (Koren et al. 2009) is a classical modelling technique that is widely used to model examinees by latent factors (Toscher and Jahrer 2010; Thai-Nghe et al. 2010; Desmarais 2012).

In spite of the importance of previous studies, there are still some limitations in existing methods. For instance, the latent factors in MF are unexplainable for describing the knowledge state of examinees' cognition. Comparatively, the results of CDMs could lead to a better interpretation. However, CDMs can only analyze examinees based on simple objective problems, and the information from subjective problems is largely underexplored (Liu and Jansen 2015). As shown in Figure 1, the objective problems differ from subjective ones in two respects: (1) Answer type, where an objective problem has a standard answer while the subjective problem has no such preset answer; (2) Scoring manner, where the response to an objective problem is objectively either correct (the same as the answer) or wrong (anything else). Comparatively, the response to a subjective problem can be totally correct, totally wrong, or partially correct based on the subjective judgement given by the reviewers (teachers). Obviously, it is hard for examinees to respond to subjective problems correctly by guessing an answer or wrongly by carelessness (e.g., a slip of the pen). Thus, these subjective problems measure the examinees much better, and it is of significant importance to extract information from both objective and subjective problems for cognitive modelling rather than simply ignoring the subjective problems or treating them as objective problems (Samejima 1972). There are several challenges: Is it possible to handle the dichotomous scores of objective problems and the polytomous scores of subjective problems simultaneously? Is it possible to achieve both precise and interpretable cognitive analysis? What is the best approach to cognitively assessing examinees based on this form of cognitive modelling (e.g., predicting examinee future performance, detecting examinee characteristics like slip and guess, and visualizing the cognitive diagnosis results)?

To conquer these challenges, in our preliminary work, we proposed a *fuzzy cognitive diagnosis framework* (FuzzyCDF) for examinees' cognitive modelling (Wu et al. 2015). FuzzyCDF is a four-tier (i.e., latent trait, skill proficiency, problem mastery, and problem score) generative model to capture the relationship between examinees' knowledge state and their performance on both objective

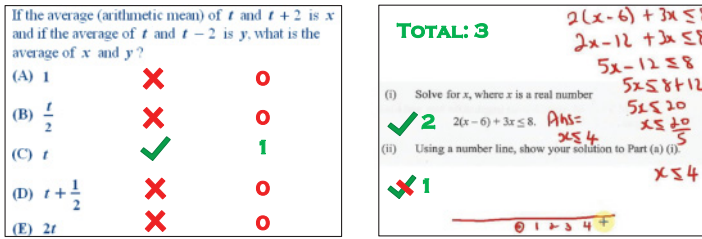


Fig. 1. An example of objective and subjective problems.

and subjective problems. Since the response to a subjective problem may be partially correct, we adopted a solution inspired by fuzzy systems. Specifically, we first fuzzified the skill proficiency of examinees from a determinative variable (mastered/non-mastered) to a fuzzy one (the degree of mastery). Then we assumed that the skill interactions on objective and subjective problems satisfy two different hypotheses: conjunctive and compensatory (Pardos et al. 2008); we fuzzified the problem mastery of examinees based on these two hypotheses by using fuzzy logic. Next, we modelled the generation of problem scores (examinee performance) by considering two exceptions: slip and guess. In this way, the whole diagnosis framework is built. We also proposed a Markov-chain Monte Carlo (MCMC) sampling algorithm to estimate the parameters, and the effectiveness of FuzzyCDF was finally verified by predicting examinee performance.

In this article, with the help of FuzzyCDF, we further address two additional cognitive assessment tasks: slip and guess detection and cognitive diagnosis visualization. Specifically, an effective Bayesian posterior-based method is proposed to calculate the probability of slip and guess given the examinees' responses. In this way, we can better understand the reasons behind the examinees' responses (for example, is a wrong response because of the carelessness/slipping of the examinee or because of her low mastery of the problem?). On the other hand, combining with the labelled Q-matrix, we can depict the knowledge structure, i.e., the inferred specific skill proficiency, of each examinee for visualized cognitive diagnosis analysis.

**Overview.** The rest of this article is organized as follows. In Section 2, we introduce the related work. Section 3 states the preliminaries and basic concepts in cognitive modelling. Section 4 details the whole framework of our FuzzyCDF. In Section 5, we specify how to accomplish three cognitive diagnosis tasks with the help of FuzzyCDF. Section 6 shows the experimental results. Finally, conclusions are given in Section 7.

## 2 RELATED WORK

In this section, we first introduce existing cognitive modelling methods from two perspectives: data-mining methods and cognitive diagnosis. Then, we discuss the related work on cognitive assessment tasks.

### 2.1 Data Mining Methods

An increasing number of studies have demonstrated the effectiveness of matrix factorization (MF) for cognitive modelling (Liu et al. 2015; Chen et al. 2017) as well as predicting examinee performance by factorizing the score matrix. For instance, Toscher and Jahrer (2010) utilized singular value decomposition (SVD) and other factor models to model examinees. In Thai-Nghe et al. (2010), the MF technique was compared with regression methods for predicting examinee performance. Thai-Nghe and Schmidt-Thieme (2015) adopted relational MF for modelling examinees in intelligent tutoring systems. In addition, Desmarais (2012) and Sun et al. (2014) applied non-negative

matrix factorization to infer the  $Q$ -matrix. In MOOCs, Díez Peláez et al. (2013) proposed MF-based methods to model learning preferences. However, the latent factors inferred by traditional MF models are usually unexplainable, i.e., each dimension of the factor vector cannot correspond to a specific skill. In this work, we will adopt cognitive diagnosis of examination data to obtain more interpretative results.

## 2.2 Cognitive Diagnosis

In educational psychology, many CDMs have been developed to mine examinees' skill proficiency (DiBello et al. 2006). CDMs can be roughly divided into two categories: continuous and discrete. The fundamental continuous CDMs are *item response theory* (IRT) models (Rasch 1961; Birnbaum 1968; Embretson and Reise 2013), which characterize examinee by a continuous variable, i.e., latent trait, and use a logistic function to model the probability that an examinee correctly solves a problem. For discrete CDMs, the basic method is *deterministic inputs, noisy "and" gate model* (DINA) (Haertel 1984; Junker and Sijtsma 2001; De La Torre 2011). DINA describes an examinee by a latent binary vector that denotes whether she has mastered the skills required by the problem, and a given  $Q$ -matrix is used to guarantee the interpretation of the diagnosis results. The DINA-based models are applied to further specific educational scenarios, such as differential item functioning assessment (Hou et al. 2014), learning team formation (Liu et al. 2016) and comprehension test validation, and difficulty estimation across multiple forms (Clark 2013; Huang et al. 2017). Though discrete CDMs are interpretable, their diagnosis results are usually not accurate enough. Furthermore, existing methods can not handle the subjective problems.

## 2.3 Cognitive Assessment Tasks

Based on the cognitive diagnosis results, e.g., the proficiency of specific skills, cognitive assessment tasks can be accomplished for deeper analysis of examinees. To discover what makes a problem easier or harder for examinees, predicting examinee performance has been viewed as a key task in Cen et al. (2006), Toscher and Jahrer (2010), Thai-Nghe et al. (2010), and De La Torre (2011) in the fields of cognitive diagnosis and data mining. Meanwhile, many efforts have been made to analyze the motivations and sentiment of examinees in problem-solving scenarios, such as slip and guess detection (Baker et al. 2008; Gu et al. 2014) and gaming-the-system (Baker et al. 2004; Wu et al. 2017). In addition, cognitive diagnosis results can be visualized for further analysis in a more convenient and interpretative way. For instance, Ben-Naim et al. (2008) proposed a solution trace graph to describe the problem-solving process, while Leony et al. (2012) provided another visualization tool to represent the learning systems. In this article, we also address three cognitive assessment tasks, i.e., predicting examinee performance, slip and guess detection, and cognitive diagnosis visualization, based on our cognitive diagnosis model.

## 3 PRELIMINARIES

In this section, we introduce the basic concepts of cognitive diagnosis, describe the characteristics of objective and subjective problems, and present the basic concepts of both fuzzy set and fuzzy logic.

### 3.1 Cognitive Diagnosis

As shown in Figure 2, examinees are usually required to participate in the pre-designed examination to assess their proficiency. Based on the examination results (in the form of score matrix with each examinee as a row and each problem as a column), psychometricians propose probabilistic models to infer the actual knowledge state of each examinee (e.g., the proficiency of one or multiple specific skills) by fitting the real data. This type of cognitive modelling is formally named *cognitive*

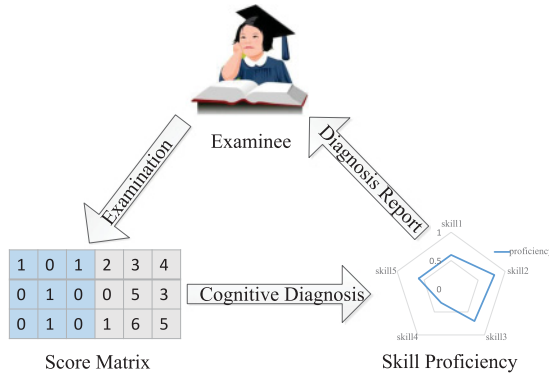


Fig. 2. A toy process of cognitive diagnosis.

	Skill 1	Skill 2	Skill 3	Skill 4
Problem 1	1	0	1	1
Problem 2	0	1	1	0
Problem 3	1	1	1	0

Fig. 3. A toy example of Q-matrix.

Table 1. Main Differences between Objective and Subjective Problems

Attribute	Objective	Subjective
Answer Type	closed, limited with standard answer	open, theoretically unlimited without standard answer
Scoring Manner	dichotomous, objectively	polytomous, subjectively
Superiority	easy for designing, well-formed for cognitive modelling	richer solving information for capturing the actual knowledge level
Inferiority	considerable effects of slip and guess	irregular response for unified modelling
Examples	chosen problem	free-response problem

diagnosis (DiBello et al. 2006). Furthermore, the obtained diagnosis results can be reformulated as a diagnosis report, which can help examinees, instructors, and educators improve the whole learning cycle. The probabilistic models to fit the score matrix and infer the knowledge state are formally named CDMs. During cognitive diagnosis, to specify the required skills (e.g., knowledge components, such as *triangles* and *inequalities*, or problem-solving abilities, like *calculation* and *induction*) of each problem, a Q-matrix (Tatsuoka 1984) is predefined by the examination designer or education experts. Figure 3 shows a toy example of a Q-matrix. For instance, an examinee is required to master Skill 2 and 3 to figure out Problem 2.

In this article, we also propose a cognitive diagnosis model to mine the actual knowledge state of examinees accurately and interpretatively, and it is based on the real-world examination data with the given Q-matrices.

### 3.2 Objective and Subjective Problems

In general, an examination consists of two different types of problems: objective and subjective problems (Liu and Jansen 2015). Table 1 summarizes the differences between these problems.

Indeed, any problem is either objective or subjective depending on the answer type and scoring manner. An objective problem has a clear correct answer, i.e., a standard answer, while a subjective

problems is anything else. With the standard answer, objective problems are judged as either correct or wrong. In contrast, the response to a subjective problem could be marked as correct, wrong, or partially correct according to the subjective judgement of a reviewer (teacher).

It is relatively easy and convenient to design an objective problem and predefine limited answer scenarios, and this further benefits the cognitive diagnosis due to the well-formed binary response. Therefore, a series of traditional psychometrical models are proposed based on objective problems (Rasch 1961; Birnbaum 1968; Embretson and Reise 2013; Haertel 1984; Junker and Sijtsma 2001; De La Torre 2011). Nevertheless, one obvious flaw with the pre-set answers of the objective problems is that it cannot determine whether the examinee gets the correct answer based on her knowledge or just a guess. Designing subjective problems differs from evaluating examinees with theoretically unlimited types of answers, and it is both harder and less intuitive for psychometricians to cognitively analyze the polytomous scores. However, the richer information in the responses (like the detailed solving process) provides a better way of capturing the actual knowledge state of examinees. For instance, it is much more difficult to answer a subjective problem correctly just by guessing.

In this article, instead of ignoring subjective problems or simply treating them as objective ones (Samejima 1972), we develop a more general cognitive diagnosis framework to handle the two types of problems simultaneously.

### 3.3 Fuzzy Set and Fuzzy Logic

In classical set theory, the membership of elements in a set is determinatively binary, and this is consistent with determinative concepts such as black versus white, success versus failure, and correct versus wrong. To handle the “fuzzy” concept, fuzzy sets (the sets whose elements have degrees of membership) are introduced (Jantzen 2013; Zadeh 1965). Therefore, fuzzy sets can address concepts such as whether a proposition is true to some degree, partially black/white, and partially correct/wrong.

A fuzzy set is a pair  $(U, \mu)$ , where  $U$  is a set and  $\mu : U \rightarrow [0, 1]$  is a membership function. For each  $x \in U$ , the value  $\mu(x)$  is called the degree of membership of  $x$  in  $(U, \mu)$ . For instance, let  $(U, \mu)$  be the fuzzy set of the black objects, and then an object  $x$  with  $\mu(x) = 1$  is in  $(U, \mu)$  (absolutely black);  $x$  with  $\mu(x) = 0$  is not in  $(U, \mu)$  (absolutely not black),  $x$  with  $0 < \mu(x) < 1$  is partially in  $(U, \mu)$  (somewhat black).

In terms of the relationship between one element and multiple fuzzy sets, researchers developed fuzzy logic as an extension of multi-valued logic (Klir and Yuan 1995). Classical multi-valued logic operations such as “AND” and “OR” can be generalized by fuzzy set operations like fuzzy intersection and fuzzy union.

In this article, we adopt fuzzy set theory to model examinees’ knowledge state and cognitive level. In this way, the specific degree of skill proficiency can be measured rather than simply classified as absolutely mastered or non-mastered. Going one step further, for modelling problem mastery, we utilize fuzzy logic to effectively capture two kinds of skill interactions to combine the fuzzified proficiency of multiple skills.

## 4 FUZZY COGNITIVE DIAGNOSIS

In this section, we will introduce our FuzzyCDF. As shown in Figure 4 (from top to bottom), our proposed method is a generation process that starts with examinees’ latent traits (e.g., a general ability of math) and then determines the examinees’ skill proficiency (e.g., the proficiency of *triangles* and *functions*); next, the examinees’ problem mastery is computed, and the observable scores are generated by considering slip and guess factors. For better illustration, Table 2 shows some

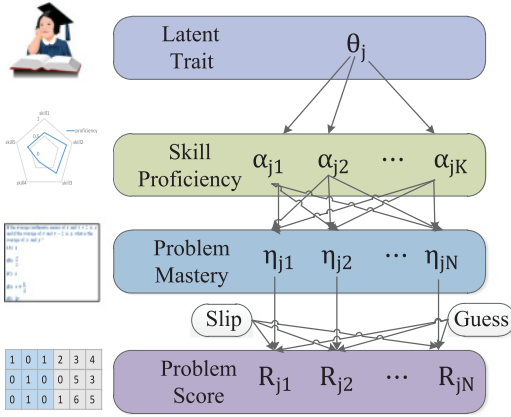


Fig. 4. Four-tier fuzzy cognitive diagnosis framework.

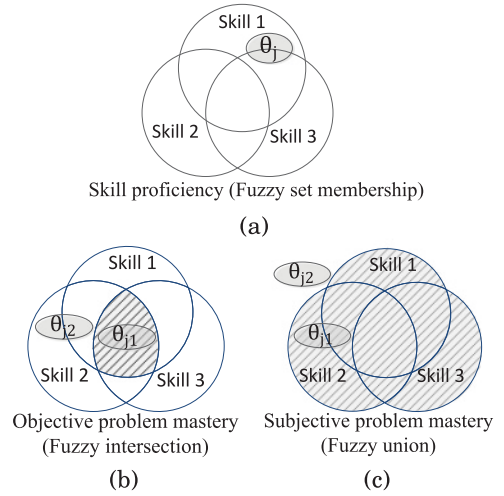


Fig. 5. Fuzzified skill proficiency and problem mastery.

Table 2. Some Important Notations

Notation	Description
$R_{ji}$	the score of examinee $j$ on problem $i$
$\theta_j$	the high-order latent trait of examinee $j$
$\alpha_{jk}$	the proficiency of examinee $j$ on skill $k$
$a_{jk}, b_{jk}$	the discrimination, difficulty of examinee $j$ on skill $k$
$\mu_k$	the membership function of fuzzy set related to skill $k$
$\eta_{ji}$	the mastery of examinee $j$ on problem $i$
$s_i, g_i$	the slip and guess factors of problem $i$

math notations, and each step of FuzzyCDF and the training procedure will be specified in the following subsections.

### 4.1 Fuzzifying Skill Proficiency

Psychologically, each examinee  $j$  has a high-order latent trait  $\theta_j$  (e.g., a general ability of math) (Rasch 1961; Birnbaum 1968). In this subsection, we show the way to get the proficiency of an examinee in specific skills (e.g., problem-solving skills like *calculation*) from her latent trait. As shown in the top two tiers of Figure 4, this is the first step in cognitive diagnosis models. Here, the challenge is that we have to handle both objective and subjective problems, simultaneously.

In DINA-based CDMs, the examinees’ skill proficiency is assumed to be determinatively mastered (i.e., 1) or nonmastered (i.e., 0), so that this modelling can fit objective problems with absolutely correct or wrong responses. However, for a subjective problem that has a partially correct response, the above “absolutely” modelling on skill proficiency cannot fit well as shown in the experiments (detailed in Section 6.2.1). To address this issue, we adapt fuzzy set theory to cognitive modelling (Zadeh 1965; Jantzen 2013; Chrysafiadi and Virvou 2015): We fuzzify the skill proficiency and the problem mastery to model examinees’ responses to objective and subjective problems. Having said that, “fuzzify” means redefining the original determinatively binary

variable (i.e., mastered or non-mastered) to a fuzzy one valued in  $[0,1]$ . We argue the assumption as follows:

**ASSUMPTION 1.** *The proficiency of an examinee on a skill is **the degree of membership of the examinee in a fuzzy set that is related to the skill.***

Here, we assume that a skill  $k$  is related to a fuzzy set  $(J, \mu_k)$ , where  $J$  is the set of examinees and  $\mu_k : J \rightarrow [0, 1]$  is the membership function. Then, for each  $j \in J$ , we define the proficiency of examinee  $j$  on skill  $k$ ,  $\alpha_{jk}$  as the degree of membership of  $j$  in  $(J, \mu_k)$ ,  $\mu_k(j)$ . Thus, if examinee  $j$  masters skill  $k$  to some extent, then the element  $j$  is a fuzzy member of the fuzzy set, i.e.,  $0 \leq \alpha_{jk} = \mu_k(j) \leq 1$ . In this way, we can fuzzify the skill proficiency of an examinee into a fuzzy variable valued in  $[0,1]$ . For instance, as shown in Figure 5(a), examinee  $j$  (represented by her latent trait  $\theta_j$ ) is fully included, fully excluded, and partially included by the fuzzy sets related to Skill 1, Skill 2, and Skill 3, respectively. It also means that examinee  $j$  has completely mastered, completely non-mastered, and partially mastered Skill 1, Skill 2, and Skill 3, respectively. Formally, following an IRT-like high-order logistic model (Rasch 1961; Birnbaum 1968; De La Torre and Douglas 2004; De La Torre 2011),  $\alpha_{jk}$  and  $\mu_k(j)$  are defined as:

$$\alpha_{jk} = \mu_k(j) = \frac{1}{1 + \exp[-1.7a_{jk}(\theta_j - b_{jk})]}. \quad (1)$$

The implication is that the proficiency of an examinee on a specific skill ( $\alpha_{jk}$ ) depends on the difference between the examinee's high-order latent trait ( $\theta_j$ ) and the properties of the skill: the difficulty ( $b_{jk}$ ) and discrimination ( $a_{jk}$ ) of skill  $k$  for examinee  $j$  (De La Torre and Douglas 2004). The coefficient 1.7 is an empirical scaling constant in logistic cognitive models to minimize the maximum difference between the normal and logistic distribution functions (Hulin et al. 1983; Camilli 1994). In this way, we can determine the proficiency of an examinee in specific skills from her latent trait.

## 4.2 Fuzzifying Problem Mastery

Based on the fuzzified skill proficiency in Section 4.1, we can further fuzzify the problem mastery of examinees.

Similarly, traditional CDMs like DINA handle problem mastery as a determinative binary variable, i.e., mastered or non-mastered. CDMs assume that problem mastery is a result of interaction of examinees' proficiency in required skills of this problem (Pardos et al. 2008). Thus, we will first discuss the skills' interaction on objective and subjective problems and then introduce a method to model the two kinds of interactions according to *fuzzy logic* and fuzzify the mastery of an examinee on a problem.

The skill's interaction on problems can be mainly categorized into *conjunctive* and *compensatory* (Pardos et al. 2008). Conjunctive means that an examinee must master all the required skills to solve a problem, while compensatory means that an examinee is able to solve a problem as long as she masters any skill required by the problem. As for examinations, an objective problem has a unique standard answer and cannot be answered correctly unless the examinee masters all the required skills without any omission. Thus, the skill's interaction on objective problems is usually assumed to be conjunctive (Pardos et al. 2008). In contrast, a subjective problem is a free-response one and the examinees can write not only the final answers but also the solution process, which includes writing the equations, deducing, calculating, and so on (a toy example is shown in the right part of Figure 1). That is, given a subjective problem and the required skills, the more skills an examinee masters, the higher the score she will achieve for this problem. Therefore, in this study,



we assume that the skill's interaction on subjective problems is compensatory. In sum, we propose an assumption about skill interaction on the problems as follows:

**ASSUMPTION 2.** *The skills' interaction on objective (or subjective) problems is **conjunctive** (or **compensatory**).*

Now, we model these two kinds of interactions in a fuzzy way to infer the problem mastery. Specifically, given the set of examinees  $J$ , suppose we have a problem  $i$  requiring Skill 1, Skill 2, and Skill 3, with their fuzzy sets  $(J, \mu_1)$ ,  $(J, \mu_2)$ , and  $(J, \mu_3)$ . As shown in Figure 5(b), examinee  $j_1$  is fully included by the fuzzy sets related to Skill 1, Skill 2, and Skill 3 and examinee  $j_2$  is not included by the fuzzy set related to Skill 3. This means that if we adopt the conjunctive assumption (similarly to the "AND" operation), examinee  $j_1$  has mastered problem  $i$  (because she has mastered all the skills needed by problem  $i$ ) and examinee  $j_2$  has not. In addition, as shown in Figure 5(c), examinee  $j_1$  is fully included by the fuzzy sets related to Skill 1 and Skill 2, and examinee  $j_2$  is not included by any of the fuzzy sets. Here, under the compensatory assumption (similarly to the "OR" operation) examinee  $j_1$  could possibly master problem  $i$ , because she has mastered at least one of the required skills. In summary, the set of the examinees who master all (any) of the skills required for problem  $i$  is the intersection (union) of the fuzzy sets related to the skills. Thus, we propose an assumption to infer the problem mastery as:

**ASSUMPTION 3.** *If the skills' interactions between each other on a problem is conjunctive (or compensatory), then the mastery of an examinee on this problem is the degree of membership of this examinee in the **intersection** (or **union**) set of the fuzzy sets related to the skills required by the problem.*

Formally, given a  $\mathbf{Q}$ -matrix with  $K$  skills, the mastery of an examinee  $j$  on an objective problem  $i$ ,  $\eta_{ji}$ , is defined as the following equation under a conjunctive assumption:

$$\eta_{ji} = \bigcap_{1 \leq k \leq K, q_{ik}=1} \mu_k(j). \quad (2)$$

Similarly,  $\eta_{ji}$  for subjective problems is defined as the following equation under a compensatory assumption:

$$\eta_{ji} = \bigcup_{1 \leq k \leq K, q_{ik}=1} \mu_k(j). \quad (3)$$

Here,  $q_{ik}$  from the  $\mathbf{Q}$ -matrix indicates whether problem  $i$  requires skill  $k$  (1 means required and 0 means non-required). Without loss of generality, we adopt the simplest fuzzy intersection and union operation<sup>1</sup> (*the standard fuzzy intersection and union* (George J and Bo 2008)) as:

$$\begin{aligned} \mu_{A \cap B}(x) &= \min(\mu_A(x), \mu_B(x)). \\ \mu_{A \cup B}(x) &= \max(\mu_A(x), \mu_B(x)). \end{aligned} \quad (4)$$

In this way, we could fuzzify the mastery of each examinee on every problem ( $\eta_{ji}$ ), whether an objective or subjective problem.

Here, we compare our fuzzy cognitive modelling with traditional methods as shown in Table 3. Note that the introduction of fuzzification actually benefits our cognitive modelling in three respects: (1) obtaining the more precise representation of skill proficiency for each examinee, (2) fitting both the objective (dichotomous score) and subjective (polytomous score) problems well, and (3) providing a convenient method (well-defined fuzzy logic) to incorporate two kinds of education hypotheses (conjunctive/compensatory) for a combination of multiple skill proficiency.

<sup>1</sup>The effects of different kinds of fuzzy set operation are out of the scope of this article.

Table 3. Comparison between Traditional and Fuzzy Cognitive Modelling

	Traditional Methods	Fuzzy Cognitive Modelling
Skill Proficiency	binary variable in $\{0,1\}$	fuzzy variable in $[0,1]$
Problem Mastery	binary variable in $\{0,1\}$	fuzzy variable in $[0,1]$
Conjunctive Interaction	“AND” operation	fuzzy intersection
Compensatory Interaction	“OR” operation	fuzzy union
Applicable Scope	objective problems (dichotomous scores)	objective and subjective problems (dichotomous or polytomous scores)

### 4.3 Generation of Examinees’ Scores

With the problem mastery defined in Section 4.2, we can now determine examinees’ scores on problems ( $R$ ). Specifically, we take two exceptions, slip and guess (Baker et al. 2008), into account and adopt two distributions to simulate the generation of scores of objective and subjective problems.

In a real-world examination, the score of an examinee on a problem depends not only on the examinee’s problem mastery. For instance, an examinee who is unable to solve the problem can get a correct response by guessing an answer (e.g., choosing “C” as the final answer somehow). Meanwhile, someone who is able to do it correctly may get a wrong response as a consequence of carelessness (e.g., a slip of the pen) (Embretson 1985). Here, we assume that each problem has its own slip and guess factors and consider these two exceptions to model the generation of examinees’ scores.

Meanwhile, we handle the different score patterns of objective and subjective problems. With either a correct or wrong response, the score of an examinee on an objective problem can be coded to a binary variable with a value in  $\{0,1\}$ . Thus, we adopt a *Bernoulli* distribution to model the scores of examinees on objective problems. Considering different score scales of subjective problems, we normalize the scores on a subjective problem by dividing the full score of the problem into a continuous variable with a value in  $[0,1]$ . Then, we assume that the score of examinees on subjective problems follow a *Gaussian* distribution, which is widely used in the literature (Mnih and Salakhutdinov 2007).

Formally, combining the problem mastery of the examinees and the exceptions of slip and guess, we simulate the generation of the scores as follows:

$$P(R_{ji} = 1 | \eta_{ji}, s_i, g_i) = (1 - s_i)\eta_{ji} + g_i(1 - \eta_{ji}). \quad (5)$$

$$P(R_{ji} | \eta_{ji}, s_i, g_i) = \mathcal{N}(R_{ji} | [(1 - s_i)\eta_{ji} + g_i(1 - \eta_{ji})], \sigma^2). \quad (6)$$

Equations (5) and (6) stand for objective problems and subjective problems, respectively.  $\mathcal{N}(\cdot | \mu, \sigma^2)$  is the probability density function of a Gaussian distribution with mean  $\mu$  and variance  $\sigma^2$ . Here,  $R_{ji}$  denotes the score (normalized score for subjective problem) of examinee  $j$  on problem  $i$ ,  $\eta_{ji}$  is the mastery of examinee  $j$  on problem  $i$  computed by Equation (2) or Equation (3),  $s_i$  and  $g_i$  denote the slip and guess factors<sup>2</sup> of problem  $i$ , and  $\sigma^2$  is the variance of the normalized score of an examinee on a subjective problem. Thus,  $(1 - s_i)\eta_{ji}$  means this examinee masters the problem and answers it successfully (i.e., without carelessness), while  $g_i(1 - \eta_{ji})$  represents that the examinee guesses a correct response without mastery. That is, these are the two ways for an examinee to give a correct response.

<sup>2</sup>Note that we model the slip and guess factor of each problem and we will address the slip and guess factor of each response in the later sections.

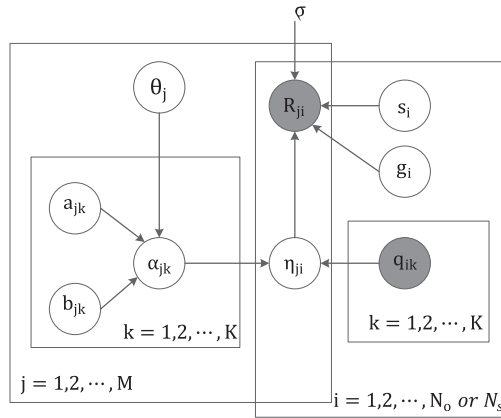


Fig. 6. The graphic model of FuzzyCDF.

**Summary.** To better understand our proposed FuzzyCDF method, we represent it using a graphic model as shown in Figure 6. Here, what we can observe are the score matrix  $\mathbf{R}$  with  $M$  examinees,  $N_o$  objective and  $N_s$  subjective problems, and the  $\mathbf{Q}$ -matrix with  $K$  skills (if problem  $i$  requires skill  $k$ , then  $q_{ik} = 1$ ). An examinee  $j$  is related to skill proficiency  $\alpha_{jk}$ ,  $k = 1, 2, \dots, K$ , which depends on high-order latent trait  $\theta_j$  and skill parameters  $a_{jk}, b_{jk}$ ,  $k = 1, 2, \dots, K$  (Equation (1)). A problem mastery  $\eta_{ji}$  is determined by required skill proficiency  $\alpha_{jk} | q_{ik} = 1$  (Equation (2) or Equation (3)), and a problem score  $R_{ji}$  is influenced by  $\eta_{ji}$  and problem parameters  $s_i, g_i$  (Equation (5) or Equation (6)). Please also note that  $\sigma$  is only used for modelling the generation of the score of an examinee on a subjective problem.

#### 4.4 Training Algorithm for FuzzyCDF

In this section, we will introduce an effective training algorithm using the MCMC method (Gilks 2005) for the proposed FuzzyCDF model, i.e., to infer the unshaded variables in Figure 6. Using the observed score matrix  $\mathbf{R}$  and the  $\mathbf{Q}$ -matrix, we could compute the full conditional probability of the parameters to be estimated. Specifically, following the settings adopted in the HO-DINA model (De La Torre and Douglas 2004), we assume the prior distributions of the parameters in FuzzyCDF as

$$\begin{aligned}
 \theta_j &\sim \mathcal{N}(\mu_\theta, \sigma_\theta^2), a_{jk} \sim \ln\mathcal{N}(\mu_a, \sigma_a^2), b_{jk} \sim \mathcal{N}(\mu_b, \sigma_b^2), \\
 s_i &\sim \text{Beta}(v_s, w_s, \min_s, \max_s), \\
 g_i &\sim \text{Beta}(v_g, w_g, \min_g, \max_g), \\
 1/\sigma^2 &\sim \Gamma(x_\sigma, y_\sigma),
 \end{aligned} \tag{7}$$

where  $\text{Beta}(v, w, \min, \max)$  is a four-parameter *Beta* distribution that has two shape parameters  $v$  and  $w$  and is supported on the range  $[\min, \max]$ . Then, the joint posterior distribution of  $\theta, \mathbf{a}, \mathbf{b}, \mathbf{s}, \mathbf{g}$ , and  $\sigma^2$  given the score matrix  $\mathbf{R}$  is

$$\begin{aligned}
 &P(\theta, \mathbf{a}, \mathbf{b}, \mathbf{s}, \mathbf{g}, \sigma^2 | \mathbf{R}) \\
 &\propto L(\mathbf{s}, \mathbf{g}, \sigma^2, \theta, \mathbf{a}, \mathbf{b})P(\theta)P(\mathbf{a})P(\mathbf{b})P(\mathbf{s})P(\mathbf{g})P(\sigma^2),
 \end{aligned} \tag{8}$$

where  $L$  is the joint likelihood function of FuzzyCDF:

$$L(\mathbf{s}, \mathbf{g}, \sigma^2, \theta, \mathbf{a}, \mathbf{b}) = L_o(\mathbf{s}, \mathbf{g}, \theta, \mathbf{a}, \mathbf{b})L_s(\mathbf{s}, \mathbf{g}, \sigma^2, \theta, \mathbf{a}, \mathbf{b}), \tag{9}$$

where  $L_o$  and  $L_s$  denote the joint likelihood functions of objective and subjective problems, respectively, and they can be defined according to Equations (5) and (6) as follows:

$$L_o(\mathbf{s}, \mathbf{g}, \theta, \mathbf{a}, \mathbf{b}) = \prod_j^M \prod_i^{N_o} (X_{ji})^{R_{ji}} (1 - X_{ji})^{1-R_{ji}}, \quad (10)$$

$$L_s(\mathbf{s}, \mathbf{g}, \sigma^2, \theta, \mathbf{a}, \mathbf{b}) = \prod_j^M \prod_i^{N_s} \mathcal{N}(R_{ji}|X_{ji}, \sigma^2), \quad (11)$$

where  $X_{ji} = (1 - s_i)\eta_{ji} + g_i(1 - \eta_{ji})$ . Note that  $\eta_{ji}$ , i.e., the problem mastery of examinee  $j$  on problem  $i$ , can be calculated given the  $\mathbf{Q}$ -matrix by using Equations (1), (2), (3), and (4). Then, the full conditional distributions of the parameters given the observed score matrix  $\mathbf{R}$  and the rest of the parameters are as follows:

$$P(\mathbf{a}, \mathbf{b}|\mathbf{R}, \theta, \mathbf{s}, \mathbf{g}, \sigma^2) \propto L(\mathbf{s}, \mathbf{g}, \sigma^2, \theta, \mathbf{a}, \mathbf{b})P(\mathbf{a})P(\mathbf{b}), \quad (12)$$

$$P(\theta|\mathbf{R}, \mathbf{a}, \mathbf{b}, \mathbf{s}, \mathbf{g}, \sigma^2) \propto L(\mathbf{s}, \mathbf{g}, \sigma^2, \theta, \mathbf{a}, \mathbf{b})P(\theta), \quad (13)$$

$$P(\mathbf{s}, \mathbf{g}|\mathbf{R}, \theta, \mathbf{a}, \mathbf{b}, \sigma^2) \propto L(\mathbf{s}, \mathbf{g}, \sigma^2, \theta, \mathbf{a}, \mathbf{b})P(\mathbf{s})P(\mathbf{g}), \quad (14)$$

$$P(\sigma^2|\mathbf{R}, \theta, \mathbf{a}, \mathbf{b}, \mathbf{s}, \mathbf{g}) \propto L_s(\mathbf{s}, \mathbf{g}, \sigma^2, \theta, \mathbf{a}, \mathbf{b})P(\sigma^2). \quad (15)$$

Finally, we propose a Metropolis-Hastings- (M-H) based MCMC algorithm (Hastings 1970) for parameter estimation by Algorithm 1. Specifically, we first randomize all the parameters as the initial values. For each iteration, we draw a uniformly random sample of each parameter within a predefined interval (specified by  $\delta_a$ ,  $\delta_b$ ,  $\delta_\theta$ ,  $\delta_s$ ,  $\delta_g$ , and  $\delta_\sigma$ ). Then, given the observable  $\mathbf{R}$  and the experts' knowledge  $\mathbf{Q}$ -matrix, we compute the full conditional probability of skill discrimination  $\mathbf{a}$ , skill difficulty  $\mathbf{b}$ , examinee latent trait  $\theta$ , problem slip factor  $\mathbf{s}$ , and guess factor  $\mathbf{g}$  and the variance of normalized scores of subjective problems  $\sigma^2$  by using Equations (9), (10), (11), (12), (13), (14), and (15). Next, the acceptance probability of samples can also be calculated based on the M-H algorithm. In this way, we could estimate the parameters after  $T$  iterations of sampling.

## 5 COGNITIVE ASSESSMENT

With the trained FuzzyCDF, we can easily discover the knowledge state and cognitive level of each examinee, e.g., the latent trait or skill proficiency. To better illustrate the effectiveness and interpretation of our cognitive modelling, in this section, we first briefly introduce three classical cognitive assessment tasks and then apply our FuzzyCDF model into these three tasks.

### 5.1 Cognitive Assessment Tasks

Traditional tests or examinations are designed for grading examinees based on their performance, namely scores, while cognitive assessment tasks are proposed for deeper and more interpretative understanding of examinees' actual ability (according to the cognitive diagnosis results). As shown in Figure 7, we introduce three classical tasks:

- **Predicting Examinee Performance** addresses the probability of an examinee correctly answering a problem she has never encountered before (Cen et al. 2006; Toscher and Jahrer 2010; Thai-Nghe et al. 2010; De La Torre 2011; Hartley and Mitrovic 2002; Corbett and Anderson 1994). Low prediction error on unseen data implies that the cognitive diagnosis model has accurately discovered what makes the problem easier/harder for examinees.

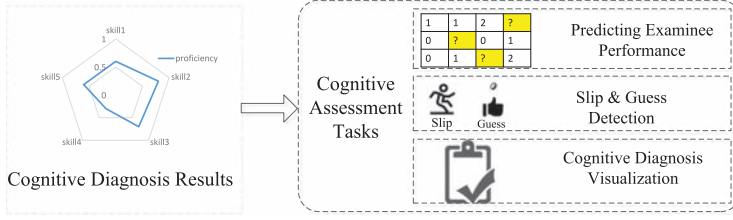


Fig. 7. Cognitive assessment tasks.

**ALGORITHM 1:** Sampling Algorithm for FuzzyCDF.**Input:** score matrix  $R$ , experts' knowledge  $Q$ -matrix**Output:** samples of each parameter

- 1: Initialize each parameter with random values
- 2: **for** each iteration  $t = 1, 2, \dots, T$  **do**
- 3: For skill discrimination  $\mathbf{a}$  and skill difficulty  $\mathbf{b}$ , draw  $\mathbf{a}_t \sim \text{Uniform}(\mathbf{a}_{t-1} - \delta_a, \mathbf{a}_{t-1} + \delta_a)$ ,  $\mathbf{b}_t \sim \text{Uniform}(\mathbf{b}_{t-1} - \delta_b, \mathbf{b}_{t-1} + \delta_b)$ , and accept  $\mathbf{a}_t, \mathbf{b}_t$  with the probability:
 
$$\min\left\{1, \frac{L(\mathbf{s}_{t-1}, \mathbf{g}_{t-1}, \sigma_{t-1}^2, \theta_{t-1}, \mathbf{a}_t, \mathbf{b}_t)P(\mathbf{a}_t)P(\mathbf{b}_t)}{L(\mathbf{s}_{t-1}, \mathbf{g}_{t-1}, \sigma_{t-1}^2, \theta_{t-1}, \mathbf{a}_{t-1}, \mathbf{b}_{t-1})P(\mathbf{a}_{t-1})P(\mathbf{b}_{t-1})}\right\}.$$
- 4: For examinee latent trait  $\theta$ , draw  $\theta_t \sim \text{Uniform}(\theta_{t-1} - \delta_\theta, \theta_{t-1} + \delta_\theta)$ , and accept  $\theta_t$  with the probability:
 
$$\min\left\{1, \frac{L(\mathbf{s}_{t-1}, \mathbf{g}_{t-1}, \sigma_{t-1}^2, \theta_t, \mathbf{a}_t, \mathbf{b}_t)P(\theta_t)}{L(\mathbf{s}_{t-1}, \mathbf{g}_{t-1}, \sigma_{t-1}^2, \theta_{t-1}, \mathbf{a}_t, \mathbf{b}_t)P(\theta_{t-1})}\right\}.$$
- 5: For problem slip factor  $\mathbf{s}$  and guess factor  $\mathbf{g}$ , draw  $\mathbf{s}_t \sim \text{Uniform}(\mathbf{s}_{t-1} - \delta_s, \mathbf{s}_{t-1} + \delta_s)$ ,  $\mathbf{g}_t \sim \text{Uniform}(\mathbf{g}_{t-1} - \delta_g, \mathbf{g}_{t-1} + \delta_g)$ , and accept  $\mathbf{s}_t, \mathbf{g}_t$  with the probability:
 
$$\min\left\{1, \frac{L(\mathbf{s}_t, \mathbf{g}_t, \sigma_{t-1}^2, \theta_t, \mathbf{a}_t, \mathbf{b}_t)P(\mathbf{s}_t)P(\mathbf{g}_t)}{L(\mathbf{s}_{t-1}, \mathbf{g}_{t-1}, \sigma_{t-1}^2, \theta_t, \mathbf{a}_t, \mathbf{b}_t)P(\mathbf{s}_{t-1})P(\mathbf{g}_{t-1})}\right\}.$$
- 6: For the variance of normalized scores of subjective problems  $\sigma^2$ , Draw  $\sigma_t^2 \sim \text{Uniform}(\sigma_{t-1}^2 - \delta_\sigma, \sigma_{t-1}^2 + \delta_\sigma)$ , and accept  $\sigma_t^2$  with the probability:
 
$$\min\left\{1, \frac{L(\mathbf{s}_t, \mathbf{g}_t, \sigma_t^2, \theta_t, \mathbf{a}_t, \mathbf{b}_t)P(\sigma_t^2)}{L(\mathbf{s}_t, \mathbf{g}_t, \sigma_{t-1}^2, \theta_t, \mathbf{a}_t, \mathbf{b}_t)P(\sigma_{t-1}^2)}\right\}.$$
- 7: **if** convergence criterion meets **then**
- 8:     **return**
- 9: **end if**
- 10: **end for**
- 11: **return**

In a further step, the analysis of examinees' performance can be used for personalized remedy/problem recommendation.

- **Slip and Guess Detection** addresses the probability that an examinee will answer a problem correctly (wrongly) as a result of a guess (slip) (Baker et al. 2008; Gu et al. 2014). It is of significant importance to mine actual learning level of examinees by eliminating slip and guess factors. Detecting slip and guess can be useful for self-evaluation and targeted remedies for examinees, learning attitude and motivation analysis for instructors, and examination design and revision for educators.
- **Cognitive Diagnosis Visualization** means specifying the strengths and shortcomings of each examinee in an interpretative way (Ben-Naim et al. 2008; Leony et al. 2012). Visualized methods or diagnosis reports are usually adopted for depicting examinees in the dimension of specific required skills as well as slip and guess factors more than examination scores. The descriptive tasks can lead to personalized learning, remedy planning, and course design.

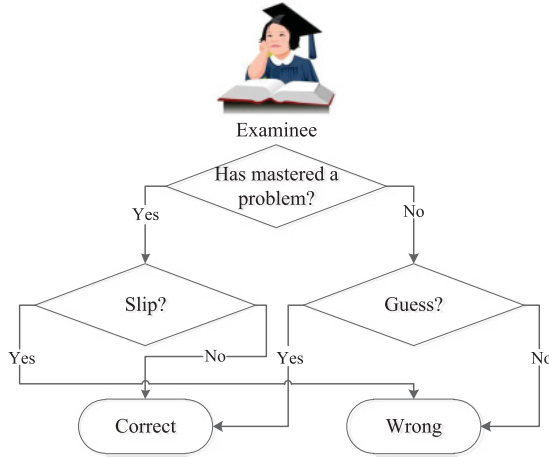


Fig. 8. The Bayesian network of slip and guess.

## 5.2 Predicting Examinee Performance

With learning history or examination records, it is of great value to discover the current knowledge state of the examinee and predict how well she will perform in the future. Usually, the effectiveness of cognitive modelling can be validated by predicting examinee performance. Here we give our solution for prediction as follows.

After the training stage mentioned in Section 4, we can easily obtain the generalized proficiency of examinees on each skill based on the estimated latent trait, skill discrimination, and skill difficulty (Equation (1)). Then, we can further compute the mastery of examinees on each problem according to the Q-matrix and problem type (objective or subjective) based on Equations (2) and (3). Combining the estimated slip and guess factors of each problem, we can now predict examinees' performance (i.e., score) on each problem based on Equations (5) and (6) as follows:

$$\hat{R}_{ji} = (1 - s_i)\eta_{ji} + g_i(1 - \eta_{ji}). \quad (16)$$

Here,  $\hat{R}_{ji}$  denotes the predicted score of examinee  $j$  on problem  $i$ . Note that the output of Equation (16) is continuous, and we can discretize them for prediction on objective problems by the following equation.

$$R_{ji}^* = \begin{cases} 1 & \text{if } \hat{R}_{ji} \geq Th_R; \\ 0 & \text{otherwise,} \end{cases} \quad (17)$$

where  $R_{ji}^*$  represents the discretized score if problem  $i$  is objective, and  $Th_R$  is the predefined threshold (usually set to be 0.5).

## 5.3 Slip and Guess Detection

In addition to providing an accurate prediction of unseen examinations, researchers are also curious about the real causes that lead to the performance of examinees. Usually, slip and guess are considered as two forms of noise in the examinations (Baker et al. 2008; Gu et al. 2014). In Section 4.3, our FuzzyCDF models the slip and guess factor of each problem by Equations (5) and (6). In this subsection, we propose an effective method for slip and guess detection of each specific response.

Cognitively, we revisit the generation of examinees' scores. As portrayed in Figure 8, a Bayesian network of slip and guess is built. Here, we divide the examinees' knowledge state on each

problem into two categories, i.e., master and nonmaster, while the responses to each problem can be classified into two scenarios, i.e., correct and wrong. An examinee mastering the problem can either respond correctly (without exceptions) or give a wrong answer (with a slip). Meanwhile, those who have not mastered the problem usually cannot get correct response unless they can score by guessing. In this way, we can summarize four types of response patterns,<sup>3</sup> i.e., wrong with a slip, wrong without a slip (which means nonmastery), correct with a guess and correct without a guess (which means mastery).

For detecting slip and guess with these four response patterns, we adopt a Bayesian posterior based solution. To be specific, the detection can be figured out by computing the posterior probability of slip or guess given the observed scores by the following Bayesian posterior equations:

$$P(s_{ji}|R_{ji}) = \frac{P(R_{ji}|s_{ji})P(s_{ji})}{P(R_{ji}|s_{ji})P(s_{ji}) + P(R_{ji}|\neg s_{ji})P(\neg s_{ji})}, \quad (18)$$

$$P(g_{ji}|R_{ji}) = \frac{P(R_{ji}|g_{ji})P(g_{ji})}{P(R_{ji}|g_{ji})P(g_{ji}) + P(R_{ji}|\neg g_{ji})P(\neg g_{ji})}. \quad (19)$$

Here,  $R_{ji}$  denotes the score of examinee  $j$  on problem  $i$ , and  $s_{ji}$  and  $g_{ji}$  denote the *slip* and *guess* of examinee  $j$  on problem  $i$ . Note that the tag  $\neg$  here means no slip or guess. With the formulated posterior, we next consider how to calculate the prior, the likelihood, and the evidence. In our cognitive framework, i.e., FuzzyCDF, we can simplify the prior computation by exploiting the estimated slip and guess parameters on each problem:

$$P(s_{ji}) = s_i, P(\neg s_{ji}) = 1 - s_i, \quad (20)$$

$$P(g_{ji}) = g_i, P(\neg g_{ji}) = 1 - g_i, \quad (21)$$

where  $s_i$  and  $g_i$  are the estimated slip and guess factors of problem  $i$ , that is, the probability that any examinee has a slip/guess on problem  $i$ . Then, considering two distributions of score generation of two types of problem, i.e., objective vs. subjective, we put forward two schemas of calculation for the likelihood and the evidence.

Specifically, examinees' scores on objective problems follow a *Bernoulli* distribution (as mentioned in Section 4.3) with either correct or wrong responses. Therefore, the likelihoods of the four types of response patterns can be defined by

$$P(R_{ji}|s_{ji}) = \eta_{ji}s_i I(R_{ji} = 0), \quad (22)$$

$$P(R_{ji}|\neg s_{ji}) = (1 - \eta_{ji})(1 - g_i) I(R_{ji} = 0), \quad (23)$$

$$P(R_{ji}|g_{ji}) = (1 - \eta_{ji})g_i I(R_{ji} = 1), \quad (24)$$

$$P(R_{ji}|\neg g_{ji}) = \eta_{ji}(1 - s_i) I(R_{ji} = 1), \quad (25)$$

where  $\eta_{ji}$  denotes the mastery of examinee  $j$  on problem  $i$ , that is, the probability that examinee  $j$  masters problem  $i$ . Therefore,  $\eta_{ji}s_i$  is the probability that examinee  $j$  masters problem  $i$  and gives a wrong response with a slip. Similarly,  $(1 - \eta_{ji})(1 - g_i)$  denotes the probability that examinee  $j$  is unable to solve problem  $i$  and can only give a wrong answer without any exceptions (e.g., guess). In contrast,  $(1 - \eta_{ji})g_i$  and  $\eta_{ji}(1 - s_i)$  represent two possible paths toward a correct response.  $I(\cdot)$  is the indicator function.

<sup>3</sup>Note that the correct answers with a slip can be viewed as another kind of "guess," and wrong responses caused by guessing are also a symbol of nonmastery. Therefore, these two responses are categorized into the patterns of "correct with a guess" and "wrong without a slip," respectively.

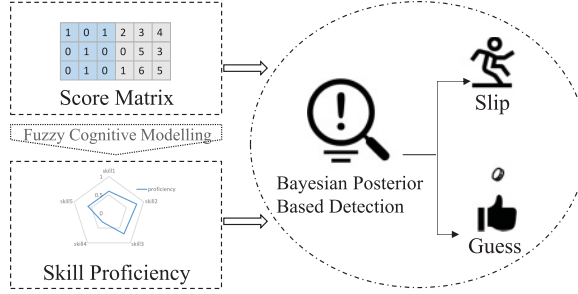


Fig. 9. Slip and guess detection.

Unlike the objective problems discussed above, each subjective problem is usually made up of several solution steps, and examinees' normalized score on subjective problems follows a *Gaussian* distribution with a value in  $[0, 1]$ . Additionally, it is much harder for an examinee to guess a response to a subjective problem (also demonstrated by the distribution of the guess parameter shown in the experimental parts) compared with objective problems. Nevertheless, considering the partially correct responses to subjective problems, we still adopt a solution of calculating the slip and guess posterior probability simultaneously. Formally, we define the likelihoods of the four types of responses by the following equations:

$$P(R_{ji}|s_{ji}) = \mathcal{N}(R_{ji}|\eta_{ji}s_i, \sigma^2)I(R_{ji} < 1), \quad (26)$$

$$P(R_{ji}|\neg s_{ji}) = \mathcal{N}(R_{ji}|(1 - \eta_{ji})(1 - g_i), \sigma^2)I(R_{ji} < 1), \quad (27)$$

$$P(R_{ji}|g_{ji}) = \mathcal{N}(R_{ji}|(1 - \eta_{ji})g_i, \sigma^2)I(R_{ji} > 0), \quad (28)$$

$$P(R_{ji}|\neg g_{ji}) = \mathcal{N}(R_{ji}|\eta_{ji}(1 - s_i), \sigma^2)I(R_{ji} > 0), \quad (29)$$

where  $\sigma^2$  is the variance of the normalized score of an examinee on a subjective problem.  $I(\cdot)$  is the indicator function.

Eventually, we can complete the slip and guess detection task as depicted in Figure 9. With the observed score matrix  $R$  and the given  $Q$ -matrix, our FuzzyCDF model can be easily trained using Algorithm 1. In the meantime, model parameters including examinees' skill proficiency, the problems' slip and guess factors and the variance of the normalized scores of subjective problems can also be estimated. Then, we can obtain the mastery of each examinee over each problem by Equations (2) and (3). Next, Equations (18) to (29) are employed to compute the prior, the likelihood, and the evidence for calculation of the Bayesian posterior of the slip and guess of each examinee on each problem.

#### 5.4 Cognitive Diagnosis Visualization

In addition to the effectiveness of a cognitive model, e.g., prediction and detection accuracy, the interpretation of cognitive models is also of great significance across research and application in the educational field (Ben-Naim et al. 2008; Leony et al. 2012). In this subsection, we focus on the interpretation of FuzzyCDF by visualizing the cognitive diagnosis results.

Unlike traditional data-mining models (e.g., matrix factorization) with latent parameters, which usually cannot describe definite meanings, cognitive diagnosis models like DINA are based on some interpretable parameters derived from psychology or pedagogy. Similarly, our FuzzyCDF is



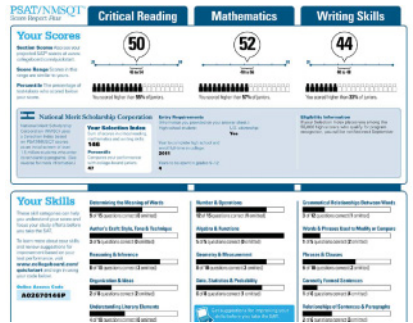


Fig. 10. An example of a score report.

Table 4. Datasets Summary

Dataset	# Examinee	# Skill	# Problem	
			Obj.	Subj.
FrcSub	536	8	20	0
Math1	4,209	11	15	5
Math2	3,911	16	16	4

guided by a Q-matrix labelled by education experts and can obtain meaningful examinee parameters (e.g., proficiency of a specific skill) and problem parameters (e.g., slip and guess factors). For instance,<sup>4</sup> as shown in Figure 10, FuzzyCDF can give a similar readable score diagnosis report for each examinee, which can point out the best and the worst learned skills. To analyze the cognitive diagnosis results, we compare and visualize the obtained skill proficiency and slip and guess factors from DINA and FuzzyCDF on some real-world examinees that are randomly sampled. The detailed comparison results are shown in the experimental part.

## 6 EXPERIMENTS

In this section, to verify the effectiveness and interpretation of cognitive modelling, we compare the performance of our FuzzyCDF against baseline approaches on three cognitive assessment tasks.

### 6.1 Experimental Setup

The experiments are conducted on three real-world datasets. The first dataset is composed of the scores of middle school students on fraction subtraction objective problems (Tatsuoka 1984; DeCarlo 2011). The other two datasets<sup>5</sup> are collected from two final mathematical exams from high school students, including both objective and subjective problems. We denote these three datasets as *FrcSub*, *Math1*, and *Math2*, respectively. Each of the datasets is represented by a score matrix and a given Q-matrix by education experts. A brief summary of these datasets is shown in Table 4. Further, Figure 11 shows a preview of these three datasets, where each column for each subfigure stands for a problem and each row above and below represents an examinee and a skill, respectively. Specifically, the three subfigures in the above show *FrcSub*'s score matrix with only dichotomous scores, *Math1*'s and *Math2*'s normalized score matrix with both dichotomous and polytomous scores; the three subfigures are actually the three Q-matrices. We can observe that

<sup>4</sup><http://collegereadiness.collegeboard.org/psat-nmsqt-psat-10/scores/student-score-reports>.

<sup>5</sup>Available at <http://staff.ustc.edu.cn/%7Eqliuql/data/math2015.rar>.

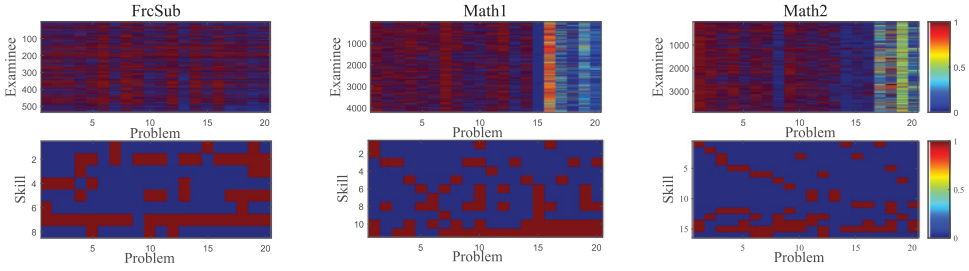


Fig. 11. The preview of the datasets.

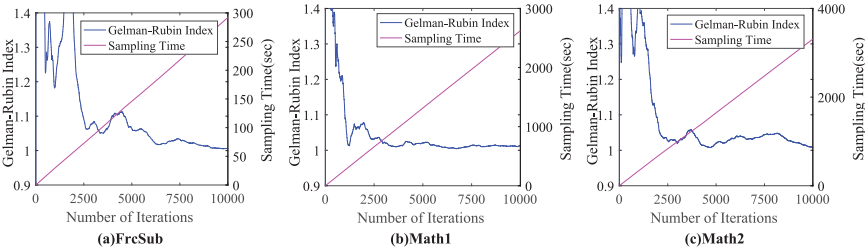


Fig. 12. Tradeoff between sampling time and convergency.

the initial problems are generally easier than the final ones and each problem is related to at least two skills.

For the prior distributions of parameters in FuzzyCDF, we follow the settings in HO-DINA (De La Torre and Douglas 2004) and tune the hyperparameters for the best performance with grid search as follows:

$$\begin{aligned} \mu_\theta &= 0, \sigma_\theta = 1; \mu_a = 0, \sigma_a = 1; \mu_b = 0, \sigma_b = 1; \\ v_s &= 1, w_s = 2, \min_s = 0, \max_s = 0.6; \\ v_g &= 1, w_g = 2, \min_g = 0, \max_g = 0.6; \\ x_\sigma &= 4, y_\sigma = 6. \end{aligned}$$

To handle the tradeoff between sampling time and convergence of the MCMC algorithm, we check *Gelman-Rubin's Rule* (Brooks and Gelman 1998) based on five parallel Markov chains, and each of the chains is run for 10,000 iterations. As shown in Figure 12, the Gelman-Runbin Index of  $\sigma^2$  (the variance of normalized scores of subjective problems) is less than 1.1 and close to 1 after 5,000 iterations, and the sampling time is linear to the number of iterations. This criterion (less than 1.2 after 5,000 iterations) is also satisfied for other structural parameters. Thus, we set the number of iterations of Algorithm 1 to 5,000 and estimate the parameters based on the last 2,500 samples to guarantee the convergence of the Markov chain. Both our FuzzyCDF and other baseline approaches are implemented by using Matlab<sup>6</sup> on a Core i5 3.1GHz machine.

<sup>6</sup>Due to the specific cognitive assessment tasks such as predicting examinee performance and slip and guess detection, we decided to implement the MCMC sampling procedure on our own. The implementation with standard Bayesian sampling software (e.g., STAN, Church) is out of scope of this work and will be addressed in the future.

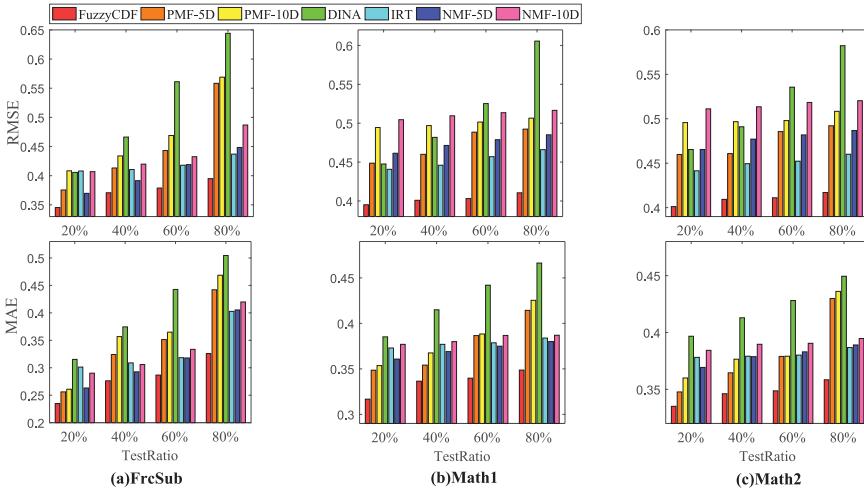


Fig. 13. Prediction task performance.

## 6.2 Cognitive Assessment Task

In this subsection, we compare the performance on three classical cognitive assessment tasks to demonstrate the effectiveness and interpretation of our proposed cognitive modelling framework.

**6.2.1 Predicting Examinee Performance (PEP) Task.** To demonstrate the effectiveness of FuzzyCDF, we conduct experiments on the PEP task, i.e., predicting the scores of the examinees over each subjective or objective problem. To observe how the methods behave at different sparsity levels, we construct different sizes of training sets, with 20%, 40%, 60%, and 80% of the score data of each examinee and the rest for testing, respectively. We use *root mean square error* (RMSE) and *mean absolute error* (MAE) as the evaluation metrics. Then, we consider baseline approaches as follows:

- *IRT*: a cognitive diagnosis method modelling examinees’ latent traits and the parameters of problems like difficulty and discrimination (Rasch 1961; Birnbaum 1968).
- *DINA*: a cognitive diagnosis method modelling examinees’ skill proficiency and the slip and guess factors of problems with a Q-matrix (Junker and Sijtsma 2001).
- *PMF*: probabilistic matrix factorization is a latent factor model projecting examinees and problems into a low-dimensional space (Mnih and Salakhutdinov 2007).
- *NMF*: non-negative matrix factorization is a latent non-negative factor model and can be viewed as a topic model (Lee and Seung 2001).

For the purpose of comparison, we record the best performance of each algorithm by tuning their parameters. Note that we treat partially correct responses as wrong for IRT and DINA due to the inapplicability of subjective problems. Figure 13 shows the PEP results of our FuzzyCDF and baseline approaches on three datasets. Here, we consider two implementations of the matrix factorization methods, PMF and NMF. That is, PMF-5D and PMF-10D (NMF-5D and NMF-10D) represent the PMF (NMF) with 5 and 10 latent factors, respectively. Thus, there are a total of seven results in each split.

From Figure 13, we observe that, over all the datasets, FuzzyCDF performs the best and the improvements are statistically significant ( $p$ -value is less than 0.001 on FuzzyCDF against other

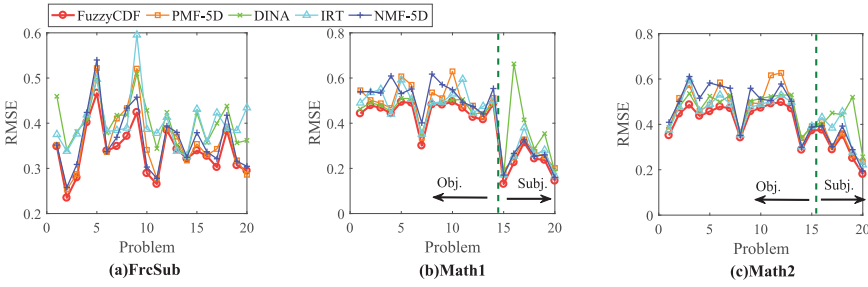


Fig. 14. Prediction performance for each problem.

Table 5. Runtime Results (Seconds)

Datasets	FuzzyCDF	PMF	DINA	IRT	NMF
FrcSub	111.26	0.14	11.19	1.77	0.07
Math1	885.63	1.27	239.56	78.87	0.344
Math2	1096.00	1.13	6184.21	68.14	0.256

baseline by using a paired  $t$ -test). Specifically, by combining educational hypotheses it beats PMF and NMF, and by quantitatively analysing examinees from a fuzzy viewpoint, it beats IRT and DINA. More importantly, with the increasing of the sparsity of the training data (training data ratio declines from 80% to 20%), the superiority of our FuzzyCDF method becomes more and more significant. For instance, when the training data is 20% (TestRatio = 80%), as shown in the sub-graphs of Figure 13(a), (b), and (c), the improvement (reducing MAE) of FuzzyCDF compared to the best baseline method IRT could reach 19%, 10%, and 8% on each datum, respectively.

In summary, FuzzyCDF captures the characteristics of examinees more precisely and it is also more suitable for the real-world scenarios, where the data are sparse and the examinees/problems are cold-start.

Fixing the training data ratio equal to 80%, Figure 14 shows the prediction performance for each specific problem in these three datasets. For simplicity, we only give the results of FuzzyCDF, and four baselines that have better performance: PMF-5D, DINA, IRT, and NMF-5D. From each sub-figure, we can observe that FuzzyCDF outperforms almost all the baselines on all the problems significantly ( $p$ -value is also less than 0.001 by using a paired  $t$ -test). Specifically, in the Math1 and Math2 datasets (the last five and four problems are subjective problems, respectively), the FuzzyCDF method can obtain the best performance for both objective and subjective problems, which in turn proves the reasonability of Assumption 2 about skill interaction on objective and subjective problems. However, matrix factorization methods (PMF and NMF) cannot fit the scores of objective problems very well, and the normal psychometric methods (IRT and DINA) are unsuitable for subjective problems. Moreover, Table 5 shows the runtime for training each method under this setting.

**6.2.2 Slip and Guess Detection.** To verify the performance of our method on the slip and guess detection task, we compare with two baseline approaches:

- *Random*: Randomly give a posterior probability, i.e., a random number in  $[0,1]$ , of slip and guess for each examinee on each problem and then obtain the top-K probable slip and guess responses by sorting the posteriors.

- *DINA*: A cognitive diagnosis method modelling examinees’ skill proficiency and the slip and guess factors of problems with a  $\mathbf{Q}$ -matrix. We take the detection strategy similar to our FuzzyCDF, i.e., compute the Bayesian posterior probability with the estimated cognitive parameters (Junker and Sijtsma 2001).

With the two baselines, we adopt two prevailing evaluation metrics, i.e., the internal validity and detection accuracy.

**Internal Validity.** In cognitive assessment, the internal validity measures the differences in observed behavior between examinees who are classified differently (Jang 2005; Roussos et al. 2006). Here, “internal” means using the training data to help verify the authenticity of the model. In this case, we assume there are four types of response patterns, i.e., wrong with or without a slip and correct with or without a guess. As discussed in Section 5.3, it is the problem mastery/nonmastery that determines the solving path toward different responses. Therefore, obvious differences exist in the observed cognitive level, say scores between examinees who have correct (wrong) responses with or without a guess (slip). Cognitively speaking, the average cognitive level or examination performance of examinees mastering the problems should manifestly differ from examinees who are nonmastering. Thus, *for each wrong (correct) response, an examinee with a higher cognitive level is assumed to have a higher slip (–guess) posterior probability*. Formally, the internal validity measure adopted in this article for slip and guess detection is referred to as the *Average Correlation (AC)* of each problem, including both objective and subjective problems, by the following equations:

$$AC_s = \frac{1}{N} \sum_{i=1}^N \rho(\mathbf{r}', \mathbf{P}_{\text{slip},i}), \quad (30)$$

$$AC_g = \frac{1}{N} \sum_{i=1}^N \rho(\mathbf{r}', \mathbf{P}_{\text{-guess},i}), \quad (31)$$

where  $\mathbf{r}' = (\bar{R}_{1\cdot}, \bar{R}_{2\cdot}, \dots, \bar{R}_{M\cdot})$  is a vector and the  $j$ th element is the average score of examinee  $j$  on each problem.  $\mathbf{P}_{\text{slip},i}$  (or  $\mathbf{P}_{\text{-guess},i}$ ) is also a vector, i.e.,  $(P(s_{1i}|R_{1i}), P(s_{2i}|R_{2i}), \dots, P(s_{Ni}|R_{Ni}))$  (or  $(P(-g_{1i}|R_{1i}), P(-g_{2i}|R_{2i}), \dots, P(-g_{Ni}|R_{Ni}))$ ), and the  $j$ th element is the slip (or –guess) posterior probability of examinee  $j$  on problem  $i$ .  $\rho$  is the correlation function, and in this work we adopt the *Pearson correlation*. Please also note that the calculation of  $AC_s$  in Equation (30) ( $AC_g$  in Equation (31)) does not cover correct (wrong) responses, since we assume there is no “correct with a slip” (no “wrong with a guess”). A larger  $AC_s$  (or  $AC_g$ ) means a more significant positive correlation between the cognitive level and the slip (–guess) posterior probability. Thus, the larger the  $AC_s$  ( $AC_g$ ), the better the internal validity.

After the training stage of each algorithm (FuzzyCDF and the baseline methods) with tuned parameters, the average internal validity (namely Average Correlation) results are shown in Figure 15. From each subfigure, we can observe that the proposed FuzzyCDF-based slip and guess detector significantly outperforms the other baseline methods, and the DINA-based method is much better than Random.

Furthermore, for better visualizing and comparing the internal validity, we rank the wrong responses of each item ordered by the related correct rate of the examinees in dataset Math1, and then, we color these responses with the detected slip posterior probability. The results are shown in Figure 16, where the higher points stand for the examinees with a bigger correct rate, while the redder and the bluer points indicate the higher and lower slip posterior probability, respectively. From the comparison, it is obvious that the FuzzyCDF-based detector actually identifies the

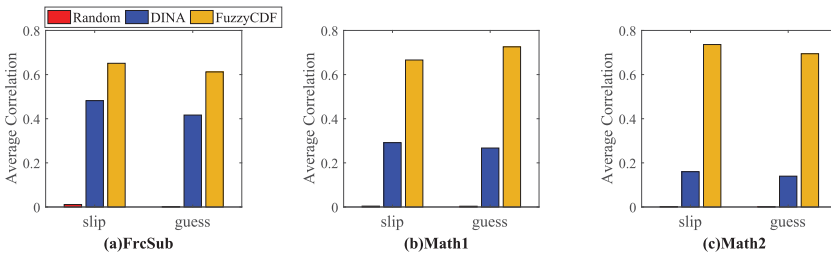


Fig. 15. Slip and guess internal validity.

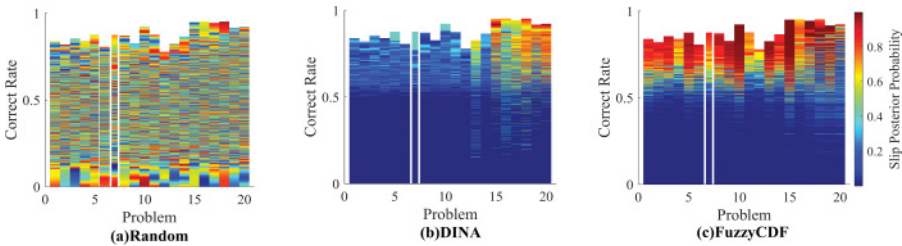


Fig. 16. Slip internal validity visualization.

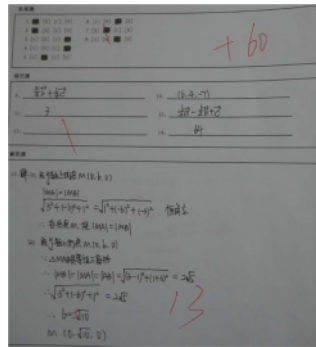


Fig. 17. A snapshot of an answer sheet.

examinees with different slip posterior probabilities better than the baselines (i.e., the DINA-based and the Random methods).

**Accuracy of Slip Detection.** Apart from internal validity, we also employ an evaluation metric with ground truth, i.e., detection accuracy. In this experiment, we randomly sample 100 examinees and scan their answer sheets for the exams from Dataset Math1.<sup>7</sup> A snapshot of an answer sheet is shown in Figure 17. Then, we ask their experienced teachers to scrutinize the sheets with the responses of subjective problems (with detailed solution steps). Note that the slip and guess of objective problems are nearly unobservable without enough evidence as the subjective ones; i.e., the ground truth of slip and guess in objective problems is basically unavailable. Finally, the teachers

<sup>7</sup>Due to the manual and time cost, we asked teachers to label only 100 examinees, and Math1 is preferred with less related skills for more convenient and precise judgement.

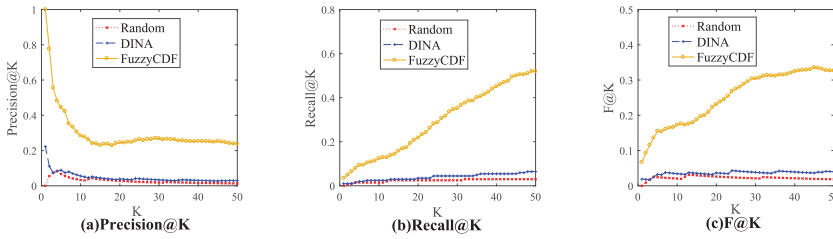


Fig. 18. Slip detection accuracy.

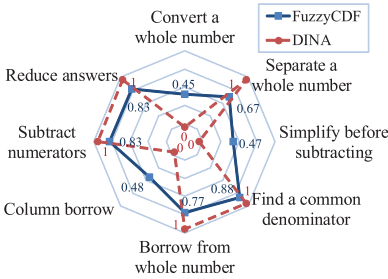


Fig. 19. Diagnosis results of an examinee on each skill.

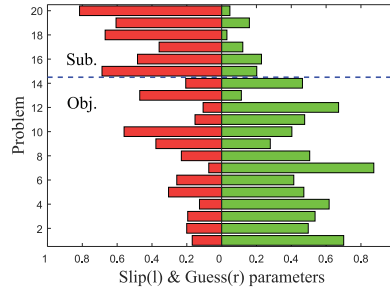


Fig. 20. Slip and guess parameters of each problem in Math1.

label the obvious slip responses<sup>8</sup> of subjective problems based on their knowledge of math, the problems, and the examinees. In total, we collect 500 labelled response data with 24 slip and 476 non-slip responses.

We adopt three widely used metrics, namely, *Precision@K*, *Recall@K*, and *F@K* as measures of accuracy evaluation of the top- $K$  ( $K = 1, 2, \dots, 50$ ) most probable slip responses (obtained by our method and the baseline approaches, respectively) in these 500 labelled responses. The results are shown in Figure 18. In terms of each metric, our FuzzyCDF-based detector is still significantly superior to the baseline methods, and the DINA-based method is slightly better than Random.

**6.2.3 Cognitive Diagnosis Visualization.** Here, we present an example of the visualized diagnosis results of an examinee on each skill in Dataset FrcSub using DINA and FuzzyCDF, respectively. The visualization results are shown in Figure 19. We can observe that both FuzzyCDF and DINA can obtain interpretatively meaningful diagnosis results with the well-designed Q-matrix. However, DINA can only distinguish whether an examinee masters a skill (1 or 0), while our FuzzyCDF can tell the extent to which the examinee masters a skill. Thus, based on our diagnosis results, an examinee can determine the true strength and shortcomings of hers. Furthermore, educators or tutoring systems can give her personalized remedy plans for improvement. Comparatively, we should note that traditional IRT and matrix factorization methods describe an examinee with latent variables, which cannot provide intuitive and interpretative results for each examinee.

We also present the slip and guess parameters of each problem in Math1, as shown in Figure 20. We can observe the obvious differences between objective and subjective problems: Generally, slip parameters of subjective problems are larger than that of objective ones, while guess parameters

<sup>8</sup>It is almost impossible to respond correctly to a subjective problem by guessing, and, meanwhile, it is also very difficult to judge whether there exists a guess only from the response.

of subjective problems are smaller than that of objective ones. The visualized results are consistent with the intuitive fact that it is much harder to guess the correct answers of subjective problems.

### 6.3 Discussion

From the experimental results of the PEP task, we can observe that FuzzyCDF outperforms the baselines on both objective and subjective problems. Slip and guess detection and cognitive diagnosis visualization demonstrated that FuzzyCDF could obtain interpretative cognitive analysis results for examinees, which can be used for composing a detailed and human-readable diagnosis report.

Nevertheless, there is still some room for improvement. First, FuzzyCDF currently suffers from the problem of high computational complexity, and we will try to design an efficient sampling algorithm in the future. Second, we can test more fuzzy set operation functions. Third, there may be some other problem types beyond objective and subjective problems, like half-open problems (Hargreaves 1984), that should be considered for cognitive modelling. In addition, we implement the MCMC sampling on our own in the current work rather than adopt a standard procedure by using existing sampling software.

## 7 CONCLUSION

In this article, we designed a fuzzy cognitive diagnosis framework, FuzzyCDF, to explore the scores of both objective and subjective problems for cognitive modelling. Specifically, we first fuzzified the skill proficiency of examinees based on a fuzzy set assumption, then fuzzified the problem mastery by mapping conjunctive and compensatory interactions into the fuzzy set operations, and next modelled the generation of the two kinds of problems with different distributions by considering slip and guess factors. Furthermore, we accomplished three cognitive assessment tasks, i.e., predicting examinee performance, slip and guess detection, and cognitive diagnosis visualization, based on our FuzzyCDF. Finally, extensive experimental results demonstrated that FuzzyCDF could quantitatively and interpretatively analyze the characteristics of each examinee and thus obtained better performance. We hope this work will lead to more studies in the future.

## REFERENCES

- Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. 2014. Engaging with massive online courses. In *Proceedings of the 23rd International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 687–698.
- Ryan Shaun Baker, Albert T. Corbett, Kenneth R. Koedinger, and Angela Z. Wagner. 2004. Off-task behavior in the cognitive tutor classroom: When students game the system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 383–390.
- Joseph E. Beck and Beverly Park Woolf. 2000. High-level student modeling with machine learning. In *International Conference on Intelligent Tutoring Systems*. Springer, 584–593.
- Dror Ben-Naim, Nadine Marcus, and Mike Bain. 2008. Visualization and analysis of student interaction in an adaptive exploratory learning environment. In *Proceedings of the International Workshop on Intelligent Support for Exploratory Environment (EC-TEL'08)*, Vol. 8.
- Allan Birnbaum. 1968. Some latent trait models and their use in inferring an examinee's ability. In *Statistical Theories of Mental Test Scores*, Frederic M. Lord and Melvin Robert Novick (Eds.). Addison-Wesley, Reading, MA.
- Stephen P. Brooks and Andrew Gelman. 1998. General methods for monitoring convergence of iterative simulations. *J. Computat. Graph. Stat.* 7, 4 (1998), 434–455.
- Hugh Burns, Carol A. Luckhardt, James W. Parlett, and Carol L. Redfield. 2014. *Intelligent Tutoring Systems: Evolutions in Design*. Psychology Press.
- Gregory Camilli. 1994. Teachers corner: Origin of the scaling constant  $d=1.7$  in item response theory. *J. Educ. Behav. Stat.* 19, 3 (1994), 293–295.
- Hao Cen, Kenneth Koedinger, and Brian Junker. 2006. Learning factors analysis—A general method for cognitive model evaluation and improvement. In *Proceedings of the International Conference on Intelligent Tutoring Systems*. Springer, 164–175.



- Yuying Chen, Qi Liu, Zhenya Huang, Le Wu, Enhong Chen, Runze Wu, Yu Su, and Guoping Hu. 2017. Tracking knowledge proficiency of students with educational priors. In *Proceedings of the 26th ACM International Conference on Information and Knowledge Management*. 989–998.
- Konstantina Chrysafiadi and Maria Virvou. 2015. Fuzzy logic for adaptive instruction in an e-learning environment for computer programming. *IEEE Transactions on Fuzzy Systems* 23, 1 (2015), 164–177.
- Amy K. Clark. 2013. Validation of a cognitive diagnostic model across multiple forms of a reading comprehension assessment. Ph.D. Dissertation. University of Kansas.
- Albert T. Corbett and John R. Anderson. 1994. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Model. User-Adapt. Interact.* 4, 4 (1994), 253–278.
- Ryan S. J. D. Baker, Albert T. Corbett, and Vincent Aleven. 2008. More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing. In *Intelligent Tutoring Systems*. Springer, 406–415.
- Jimmy De La Torre. 2011. The generalized DINA model framework. *Psychometrika* 76, 2 (2011), 179–199.
- Jimmy De La Torre and Jeffrey A. Douglas. 2004. Higher-order latent trait models for cognitive diagnosis. *Psychometrika* 69, 3 (2004), 333–353.
- Lawrence T. DeCarlo. 2011. On the analysis of fraction subtraction data: The DINA model, classification, latent class sizes, and the Q-matrix. *Applied Psychological Measurement* 35, 1 (2011), 8–26.
- Michel C. Desmarais. 2012. Mapping question items to skills with non-negative matrix factorization. *ACM SIGKDD Explor. Newslett.* 13, 2 (2012), 30–36.
- Louis V. DiBello, Louis A. Roussos, and William Stout. 2006. Review of cognitively diagnostic assessment and a summary of psychometric models. In *Handbook of statistics: Psychometrics*, Calyampudi Radhakrishna Rao and Sinharay Sandip (Eds.). Vol. 26. Elsevier, 979–1030.
- Jorge Díez Peláez, Óscar Luaces Rodríguez, Amparo Alonso Betanzos, Alicia Troncoso, and Antonio Bahamonde Rionda. 2013. Peer assessment in MOOCs using preference learning via matrix factorization. In *Proceedings of the NIPS Workshop on Data Driven Education*.
- Susan E. Embretson. 1985. Multicomponent latent trait models for test design. *Test Design: Developments in Psychology and Psychometrics* (1985), 195–218.
- Susan E. Embretson and Steven P. Reise. 2013. *Item Response Theory for Psychologists*. Psychology Press.
- George Klir and Bo Yuan. 1995. Fuzzy sets and fuzzy logic: Theory and applications. Englewood Cliffs, NJ: Prentice-Hall.
- Walter R. Gilks. 2005. Markov chain Monte Carlo. *Encyclopedia of Biostatistics*.
- Junjie Gu, Yutao Wang, and Neil T. Heffernan. 2014. Personalizing knowledge tracing: Should we individualize slip, guess, prior or learn rate? In *Proceedings of the International Conference on Intelligent Tutoring Systems*. Springer, 647–648.
- Edward Haertel. 1984. An application of latent class models to assessment data. *Appl. Psychol. Measure.* 8, 3 (1984), 333–346.
- David H. Hargreaves. 1984. Teachers questions: Open, closed and half-open. *Educ. Res.* 26, 1 (1984), 46–51.
- Danita Hartley and Antonija Mitrovic. 2002. Supporting learning by opening the student model. In *Proceedings of the International Conference on Intelligent Tutoring Systems*. Springer, 453–462.
- W. Keith Hastings. 1970. Monte carlo sampling methods using markov chains and their applications. *Biometrika* 57, 1 (1970), 97–109.
- Likun Hou, Jimmy de la Torre, and Ratna Nandakumar. 2014. Differential item functioning assessment in cognitive diagnostic modeling: Application of the wald test to investigate DIF in the DINA model. *J. Educ. Measure.* 51, 1 (2014), 98–125.
- Zhenya Huang, Qi Liu, Enhong Chen, Hongke Zhao, Mingyong Gao, Si Wei, Yu Su, and Guoping Hu. 2017. Question difficulty prediction for reading problems in standard tests. In *AAAI*. 1352–1359.
- Charles Lee Hulin, Fritz Drasgow, and Charles K. Parsons. 1983. *Item Response Theory: Application to Psychological Measurement*. Dorsey Press.
- Eunice Eunhee Jang. 2005. *A Validity Narrative: Effects of Reading Skills Diagnosis on Teaching and Learning in the Context of NG TOEFL*. Ph.D. dissertation. University of Illinois at Urbana-Champaign.
- Jan Jantzen. 2013. *Foundations of Fuzzy Control: A Practical Approach*. John Wiley & Sons.
- Brian W. Junker and Klaas Sijtsma. 2001. Cognitive assessment models with few assumptions, and connections with non-parametric item response theory. *Appl. Psychol. Measure.* 25, 3 (2001), 258–272.
- Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *Computer* 8 (2009), 30–37.
- Daniel D. Lee and H. Sebastian Seung. 2001. Algorithms for non-negative matrix factorization. In *Advances in NIPS*. 556–562.
- Derick Leony, Abelardo Pardo, Luis de la Fuente Valentin, David Sánchez de Castro, and Carlos Delgado Kloos. 2012. GLASS: A learning analytics visualization tool. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*. ACM, 162–163.

- Qi Liu, Xianyu Zeng, Chuanren Liu, Hengshu Zhu, Enhong Chen, Hui Xiong, and Xing Xie. 2015. Mining indecisiveness in customer behaviors. In *Proceedings of the IEEE International Conference on Data Mining*. 281–290.
- Yuping Liu, Qi Liu, Runze Wu, Enhong Chen, Yu Su, Zhigang Chen, and Guoping Hu. 2016. Collaborative learning team formation: A cognitive modeling perspective. In *Proceedings of the International Conference on Database Systems for Advanced Applications*. Springer, 383–400.
- Zhe Liu and Bernard J. Jansen. 2015. Subjective versus objective questions: Perception of question subjectivity in social Q&A. In *Proceedings of the International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*. Springer, 131–140.
- Yiming Ma, Bing Liu, Ching Kian Wong, Philip S. Yu, and Shuik Ming Lee. 2000. Targeting the right students using data mining. In *Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 457–464.
- Andriy Mnih and Ruslan Salakhutdinov. 2007. Probabilistic matrix factorization. In *Advances in NIPS*. 1257–1264.
- Paul D. Nichols, Susan F. Chipman, and Robert L. Brennan. 2012. *Cognitively Diagnostic Assessment*. Routledge.
- Zachary A. Pardos, Joseph E. Beck, Carolina Ruiz, and Neil T. Heffernan. 2008. The composition effect: Conjunctive or compensatory? An analysis of multi-skill math questions in ITS. In *Proceedings of the 1st International Conference on Educational Data Mining*. Citeseer.
- Georg Rasch. 1961. On general laws and the meaning of measurement in psychology. In *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 4. University of California Press, Berkeley, CA, 321–333.
- L. Roussos, L. DiBello, W. Stout, S. M. Hartz, R. A. Henson, and J. L. Templin. 2006. Diagnostic skills-based testing using the fusion-model-based arpeggio system. In *Cognitively Diagnostic Assessment in Education: Theory and Practice*, Jacqueline Leighton and Gierl Mark (Eds.). Cambridge University Press.
- Fumiko Samejima. 1972. A general model for free-response data. In *Psychometrika Monograph*. 18, 34 (4, Pt. 2).
- Yuan Sun, Shiwei Ye, Shunya Inoue, and Yi Sun. 2014. Alternating recursive method for Q-matrix learning. In *Proceedings of the 7th Int'l Conf. on Educational Data Mining*. 14–20.
- K. K. Tatsuoka. 1984. Analysis of errors in fraction addition and subtraction problems (final report for grant no. NIE-G-81-0002). University of Illinois–Urbana. *Computer-based Education Research Laboratory* (1984).
- Nguyen Thai-Nghe, Lucas Drumond, Artus Krohn-Grimberghe, and Lars Schmidt-Thieme. 2010. Recommender system for predicting student performance. *Proc. Comput. Sci.* 1, 2 (2010), 2811–2819.
- Nguyen Thai-Nghe and Lars Schmidt-Thieme. 2015. Multi-relational factorization models for student modeling in intelligent tutoring systems. In *Proceedings of the 7th International Conference on Knowledge and Systems Engineering (KSE'15)*. IEEE, 61–66.
- A. Toscher and Michael Jahrer. 2010. Collaborative filtering applied to educational data mining. In *Proceedings of the KDD 2010 Cup 2010 Workshop: Knowledge Discovery in Educational Data*. 13–23.
- Runze Wu, Qi Liu, Yuping Liu, Enhong Chen, Yu Su, Zhigang Chen, and Guoping Hu. 2015. Cognitive modelling for predicting examinee performance. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence*. AAAI Press, 1017–1024.
- Runze Wu, Guandong Xu, Enhong Chen, Qi Liu, and Wan Ng. 2017. Knowledge or gaming?: Cognitive modelling based on multiple-attempt response. In *Proceedings of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 321–329.
- Lotfi A. Zadeh. 1965. Fuzzy sets. *Inf. Control* 8, 3 (1965), 338–353.

Received December 2016; revised October 2017; accepted November 2017