scientific reports

OPEN

Cognitive analysis and path construction of Chinese students' mathematics cognitive process based on CDA

Xiaopeng Wu[1,2✉], Nanxin Li[3], Rongxiu Wu[4] & Hao Liu[5]

As a new generation of assessment theory, cognitive diagnosis can provide students with better personalized and formative learning. As an international large-scale test, the *Trends in International Mathematics and Science Study* (TIMSS) has already had some research foundation in cognitive diagnostic assessment (CDA), however, Mainland China has not yet participated. Therefore, this study aims to understand the specific performance of Mainland Chinese students within the TIMSS framework, compare their performance to other high-performing countries, and analyze the students' learning paths and progressions. With CDA techniques, this study first identified eight attributes in mathematics cognition and formed a Q-matrix based on TIMSS-2015 items to construct the diagnosis assessment. A total of 4,733 Grade 8 students from Gansu, Guangdong, Guizhou, and Shanghai in Mainland China were measured using a mixed model from the G-DINA package in software R. The findings revealed that Grade 8 students in Mainland China exhibited an absolute advantage in the mastery of mathematical cognitive processes, particularly in the traditional domains such as *Calculation and Measurement* (CM), *Operation and Solution* (OS), and *Representation Modeling* (RM). Furthermore, analyses of the second and third learning progressions demonstrated diverse knowledge states among students. Last, students with the same overall score showed substantial differences in their mastery of specific cognitive processes. This study thoroughly discusses the construction of methods for TIMSS cognitive diagnostic assessment as well as the construction of learning paths and progressions in the assessment. It also highlights the potential of CDA for assessing cognitive abilities and constructing in-depth data mining, advancing the understanding of students' cognitive strengths and weaknesses in mathematics education.

In the 1970s, *Classical test theory* (CTT), which relies on raw scores, dominated the field of measurement research. However, CTT is heavily dependent on both the test items and the participants, with scores varying across different participants and items. *Item response theory* (IRT) emerged as an alternative, offering the ability to estimate participants' ability on a common scale. From the 1970s to the 1980s, a substantial number of studies applied IRT to practical evaluation problems[1]. Despite its advantages over CTT, IRT remains a summary assessment, only capable of reporting participants' abilities in a single dimension. Critically, the IRT model "has little connection with the basic process, strategy and knowledge structure of project solving concerned by cognitive theory"[2]. Neither CTT nor IRT can reflect the psychological characteristics or cognitive processes involved in participants' responses to test items, nor can they capture participants' mastery of specific, fine-grained knowledge points. To address these limitations, Cognitive Diagnostic Modeling (CDM) was developed. CDMs overcome the shortcomings of traditional theories and align with practical educational requirements. CDMs are "designed to detect a student's specific knowledge structure or operational skills in a certain area, thereby providing detailed diagnostic information about the student's cognitive strengths and weaknesses[3]. " It presents test questions to students in the form of items, with the student's latent traits serving as assessment

[1]Faculty of Education, Northeast Normal University, Changchun, China. [2]Faculty of Education, The University of Hong Kong, Hong Kong, China. [3]Ningxia Yinchuan Experimental Primary School, Yinchuan, China. [4]Science Education Department, Harvard-Smithsonian Center for Astrophysics, Harvard University, Cambridge, USA. [5]Collaborative Innovation Center of Assessment toward Basic Education Quality, Beijing Normal University, Beijing, China. ✉email: wuxp722@nenu.edu.cn

attributes. Based on the students' responses, psychological assessment models are used to infer students' mastery of various attributes, laying the groundwork for personalized learning[4]. In essence, cognitive diagnostic assessment (CDA) is an evaluation method that applies cognitive diagnostic theory and models to analyze test data and generate detailed diagnostic insights[5].

In the mathematics education, it is suggested that mathematics teachers, mathematicians, psychometricians, and educational statisticians cooperate in research projects to recognize the potential value of additional conceptual discussion and secondary analysis directly applicable to the existing school system[6]. At present, although *Trends in International Mathematics and Science Study* (TIMSS) was not originally developed for diagnostic purposes, its items using in non-diagnostic assessment can still be applied for diagnostic purposes[7]. The structure of CDMs within the TIMSS assessment is clear and suitable for CDA, which allows for the transformation of the non-diagnostic TIMSS test into a cognitive diagnostic assessment tool with diagnostic capabilities[8]. Mathematical cognitive processes refer to students' understanding and operational processes involving mathematical knowledge and skills. As one of the two core areas assessed in the TIMSS test, this domain has garnered significant attention. TIMSS categorizes the assessment of cognitive domains into three major components: knowing, applying, and reasoning, proving a comprehensive framework for evaluating students' mathematical cognition.

There have been several cognitive diagnostic studies utilizing TIMSS data to explore students' mathematical performance. Tatsuoka and others used the rule space model (RSM) to sample and analyze the mathematical performance of Grade 8 students in TIMSS in 1999[9]. Dogan et al. used RSM to study the mathematical performance of Turkish students on TIMSS-R. A Q-matrix containing 23 attributes was used to compare the distribution of attribute mastery between Turkish and American students[10]. The results showed that Turkish students have poor ability to master attributes such as P10 (quantitative reading), S4 (approximation / estimation), S6 (mode and relationship) and S10 (solving open problems) compared with those for American students[10]. Eunkyoung et al. conducted a similar study to compare the mastery of students' attributes in South Korea, the Czech Republic and the United States[11]. Using the TIMSS-1999 dataset, Birenbaum and others also compared the attribute mastery of students in the United States, Singapore and Israel[12]. Using the similar same-year dataset, Chen conducted a CDA with three specific analyses, including the calculation of classification rate, multiple regression analysis and the comparison of attribute mastery probability in four pamphlets[4]. While most prior research focused on cross-national comparisons of students' knowledge mastery, they rarely provided individual-level analyses. Among the few studies that did, the analyses were limited to simple descriptive statistics, with no application of rigorous inferential statistical methods. This gap underscores the need for further research to provide deeper insights into individual-level cognitive diagnostic analyses.

TIMSS, as the largest international educational assessment, involves over 60 countries globally. Its purpose is to monitor achievement trends in mathematics and science among Grade 4 and 8 students from participating countries worldwide, as well as to assess curriculum implementation and identify promising teaching practices. It provides cross-national comparative data that informs educational policies. Regrettably, Mainland China, despite being the most populous nation, has not yet participated in this assessment. Therefore, by obtaining our own data through CDA, we aim to explore how Mainland Chinese students would perform under the TIMSS framework. We seek to identify differences between the results of China and those countries that perform exceptionally well and to further analyze the pathways and progressions in students' learning.

Methods
Attribute
Tatsuoka believed that attributes are production rules, project types, program operations or more general cognitive tasks. In this study, cognitive attributes were defined as the cognitive processes necessary for students to complete the test items[13]. The classification of these attributes is a constructed ordered classification according to the order of students' cognitive development starting from students' cognition when completing the test task. As one of the two core contents of it, cognition has received much attention in this field. TIMSS-2015 divided the examination of cognitive field into three parts: knowing, applying, and reasoning. Based on the three dimensions of TIMSS evaluation, seven specific attributes were formed (Table 1).

Q-matrix

The Q-matrix is a relational table used to represent how test items examine specific attributes, where 0 indicates that the attribute is not examined, and 1 indicates that the attribute is examined. In this study, a set of publicly available TIMSS-2015 questionnaires was selected as the evaluation items, and twenty experts were selected as the project calibration expert group, including eight doctoral students majoring in mathematics education, six front-line mathematics teachers in middle school, two provincial famous teachers selected by various provinces, and four experts in mathematics education in colleges and universities. They calibrated the attributes examined by the test items, for example, item 6:

Code	Attribute	Content definition
RR	Recollection and recognition	This attribute includes three parts: memory, recognition and sorting. Memorizing definitions, terms, number operations, units of measure, geometric properties and symbols; Identification of expressions, quantities and shapes, mathematical equivalence relations; Sort numbers, expressions, quantities, and shapes by common attributes, sort quantities that have a size relationship, and etc.
CM	Calculation and measurement	The attribute includes four mixed operations on natural numbers, integers, fractions, decimals and their combinations, and can carry out simple algebra and operations; Obtain useful information from diagrams, tables, texts and other resources; Be able to use measuring tools and select appropriate units of measurement
OS	Operation and solution	Conventional problems of existing solutions are solved through operations such as operations. This attribute is to use existing methods to solve simple mathematical or practical problems, often with only a single process
RM	Representation modeling	Representing data in different forms; Create equations, inequalities, and geometric figures to model practical problems; Provide solutions to practical problems
PI	Process implementation	Implement strategies and operations to solve multi-step problems related to mathematical concepts and processes, usually involving multi-step operations that reflect the procedural nature of problem solving
AE	Analysis and evaluation	Analysis of the relationship between mathematical elements such as quantity, expression and geometry; Evaluate problem solving strategies and solutions; Make valid inferences based on information and evidence. This attribute often involves new problem situations that need to be explored to solve the problem
GP	Generalization and Proof	Obtain the main information from the specific problem situation and mathematical situation, summarize mathematical relations, and express the relations in a more general form; A process by which simple existing conclusions can be proved, providing mathematical arguments to support a strategy or solution

Table 1. Attribute definitions of cognitive process.
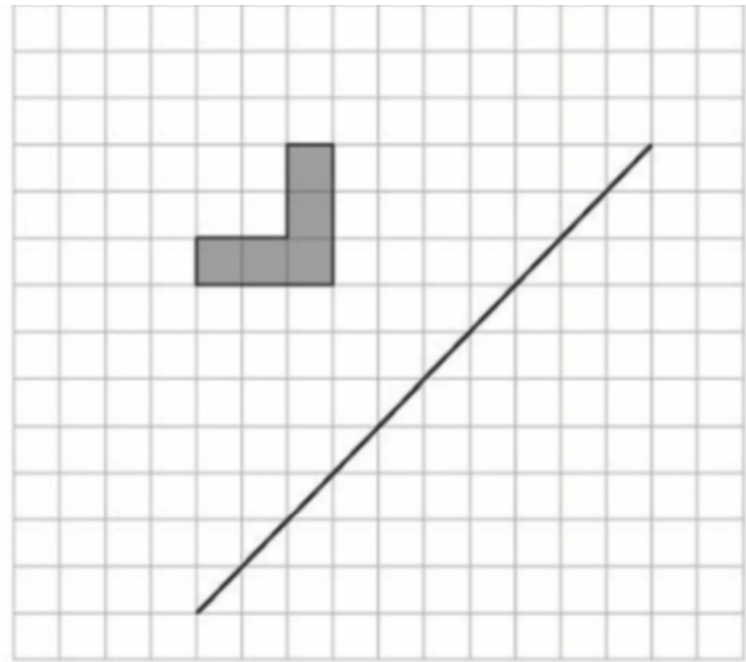
the symmetric image of a shaded image with respect to a line.

The calibration result for this item is RM (Representation Modeling).

Twenty experts coded independently and the final calibration results are presented as follows (Table 2).

According to Table 2, except that attribute OS and GP had only three items to examine, other attributes had four or more than four items to examine.

Samples

Since Mainland China has not participated in the TIMSS test, this study conducted an independent assessment involving 4,733 Grade 8 students from Gansu, Guizhou, Guangdong, and Shanghai. These regions were deliberately chosen to ensure the representativeness and diversity of the sample, reflecting the significant economic and educational disparities between China's central-western and eastern regions. Gansu and Guizhou represent less economically developed areas in the northwest and southwest, respectively, while Guangdong and Shanghai were selected as examples of more economically advanced regions in the southeast and along the eastern coast.

The assessment utilized a set of 28 items from TIMSS-2015 to evaluate the mathematical cognitive processes of these students. Additionally, to provide a broader context and comparative insights, secondary data from TIMSS assessments of high-performing regions—Singapore, Japan, South Korea, Taiwan, and Hong Kong—were analyzed. This approach allowed for an in-depth examination of Mainland Chinese students' performance in relation to international benchmarks.

Item code	RR	CM	OS	RM	PI	AE	GP
Item1	0	1	0	0	0	0	0
Item2	0	0	0	1	0	0	0
Item3	0	0	0	0	1	0	0
Item4	0	0	0	0	1	0	0
Item5	1	0	0	0	0	0	0
Item6	0	0	1	0	0	0	0
Item7	0	0	0	1	0	0	0
Item8	0	0	0	0	0	1	0
Item9	0	0	0	0	0	1	0
Item10	0	0	0	1	1	0	0
Item11	1	0	0	0	0	0	0
Item12	0	1	0	0	0	0	0
Item13	0	1	0	0	0	0	0
Item14	1	0	0	0	0	0	0
Item15	0	0	0	1	0	0	0
Item16	0	0	0	0	0	0	1
Item17	1	1	0	0	0	0	0
Item18	0	1	0	0	0	0	0
Item19	0	1	0	0	0	0	0
Item20	0	0	0	1	0	0	1
Item21	0	1	0	0	0	0	0
Item22	0	0	0	0	1	0	0
Item23	0	0	0	0	0	1	1
Item24	0	0	0	0	0	1	0
Item25	1	0	0	0	0	0	0
Item26	0	0	1	0	0	0	0
Item27	0	0	1	0	0	0	0
Item28	0	0	0	0	1	0	0

Table 2. Q-matrix of TIMSS cognitive process. Recollection and recognition (RR), Calculation and measurement (CM), Operation and solution (OS), Representation modeling (RM), Process implementation (PI), Analysis and evaluation (AE), Generalization and proof (GP).

Model selection

The key to the correct diagnosis is to choose the appropriate diagnosis model based on different cognitive assessment assumptions[5,14,15]. To select the best fitting model, this study evaluated the parameters of seven commonly used models, and selected the model with the best fitting effect through model comparisons. The seven models specifically are Deterministic Input; Noisy 'And' Gate (DINA)[16], which assumes that for a test-taker to correctly answer a particular item, they must master all the attributes required by the item. Lacking even one attribute significantly lowers the probability of correctly answering the item, making this model a non-compensatory cognitive diagnostic model; Deterministic Input; Noisy 'Or' Gate (DINO)[17], which assumes that the more attributes a test-taker masters, the higher the probability of correctly answering the item, as the attributes have an additive effect. Therefore, this model is a compensatory cognitive diagnostic model that allows attributes to compensate for each other; Reduced reparametrized unified model (RRUM)[18] introduces a penalty parameter. If a test-taker has not mastered a specific attribute, the penalty parameter reduces their probability of correctly answering the item, making the model partially compensatory. Additive Cognitive Diagnostic Model (ACDM)[19], assumes that mastering an attribute increases the probability of answering the item correctly in a linear additive fashion. This model exhibits partial compensatory effects, as attributes contribute additively to the probability of success; Loglinear CDM (LCDM)[20], is a saturated cognitive diagnostic model that integrates the categorical latent variable approach of cognitive diagnostic models with item response theory (IRT), offering flexibility and comprehensive modeling of item responses; Linear logit Model (LLM)[21] is a linear logistic regression model with cognitive diagnostic capabilities. It extends traditional linear models to accommodate cognitive diagnostic functionalities. Finally, the Mixtures Model[22] selects the optimal simplified model for each item by evaluating a combination of models, enabling a tailored approach to cognitive diagnosis.

The evaluation process is implemented using G-DINA package in software R, and the evaluation results are shown in Table 3 below:

According to Table 3, the results of model comparison show that the AIC and BIC of the mixed model are smaller than those of other models. This shows that the mixed model shows a better fitting for the data in this study.

Model	DINA	DINO	RRUM	ACDM	LCDM	LLM	Mixed model
n	183	183	187	187	191	187	186
Deviation	14410.35	14415.15	14386.50	14381.19	14373.79	14380.48	14376.57
AIC	14776.35	14781.15	14760.50	14755.19	14755.79	14754.48	**14748.57**
BIC	15572.56	15577.36	15574.11	155568.81	15586.81	15568.09	**15557.84**

Table 3. Parameter statistics of different models of TIMSS cognitive process.

Inspection of analytical tools
Item fitting analysis
The fitting effect of each item and model in the test tool is an important factor in the evaluation of cognitive diagnosis. Some studies have shown that the fitting effect of CDM and test items directly determines the accuracy of the diagnosis effect of the model[23]. The error of the square test is considered as the deviation of the potential response under the error root approximation test. The RMSEA calculation formula of project j is[24]:

$$\text{RMSEA}_j = \sqrt{\sum_k \sum_c \pi(\theta_c)\left(P_j(\theta_c) - \frac{n_{jkc}}{N_{jc}}\right)^2}$$

$\pi(\theta_c)$ represents the classification probability of the level of class C potential trait, P_j represents the probability estimated by the project response function. n_{jkc} refers to the expected number of people in dimension k of category C potential trait level in item j, N_{jc} refers to the expected number of potential trait levels of category C.

The closer the value of RMSEA is to 0, the smaller the fitting deviation is, and the better the fitting is. In the study of Oliveri and von Davier, the critical value of RMSEA is set to 0.1. When RMSEA > 0.1, it indicates that the project fitting is poor[25]. According to this standard, it can be concluded that in the cognitive process attribute, only the fitting effects for item16 and Item21 were poor (RMSEAs > 0.1), and the other overall fitting effects are acceptable.

Absolute fitting analysis
The process of model comparison is essentially a relative model-fitting evaluation. To comprehensively assess model performance, conducting an absolute fit analysis for each model is equally important. In the context of education evaluation, absolute fit analysis holds particular importance as it focuses on how well the model fits the observed response data independently, without comparing it to other models[26]. The absolute fitting index in this study adopts the Limited information of the Root Mean Square Error of Approximation (RMSEA2)[27]. which is an absolute fitting index commonly used in CDM_S. In the construction of index model, RMSEA2 is different from conventional RMSEA because it only uses two moments: univariate and bivariate interaction[28]. And the formula is as follows:

$$\text{RMSEA2}_j = \sqrt{\sum_k \sum_c \pi(\alpha_c)\left(P_j(\alpha_c) - \frac{\widehat{n}_{jkc}}{\widehat{N}_{jc}}\right)^2}$$

Where k represents the attribute, c represents a potential class of the specific attribute combination α_c, $\pi(\alpha_c)$ is the potential class α_c being evaluated, P_j is the evaluated response function, $\widehat{n}_{jkc}$ is the expected number of students of item j in category k possessing α_c, $\widehat{N}_{jc}$ is the expected number of students possessing α_c on project j[29]. The mean value of RMSEA2 is the average value of RMSEA2 of all projects, which can represent the overall fitting of the model. At present, there is no unified standard on RMSEA2 in CDMs. Some studies believe that generally RMSEA2 < 0.089 is a sufficient fit, and RMSEA2 < 0.05 is a better fit in the multidimensional item response theory[28]. Hu and other scholars believe that RMSEA2 < 0.05 in CDM_S is the standard for model fitting[30]. Through the operation of G-DINA package in R, the attribute in cognitive process is obtained, RMSEA2 = 0.0299 < 0.05, so the mixed model absolutely fits the data.

Reliability analysis
The reliability of CDA can be investigated from two perspectives: first, by treating the test as a traditional assessment and calculating the alpha coefficient under CTT; second, by assessing the test-retest consistency. Templin et al. obtained the index by calculating the correlation of attribute mastery probability of the same participants in two successive measurements under the assumption that the attribute probability mastered by the participants remains unchanged[31]. In this study, the CDA platform (flexCDMs) developed by the team of Tu Dongbo was used to evaluate the reliability[32]. Through the evaluation of the data α = 0.9079 > 0.7, which has high reliability under the CTT theory. The statistics of Templin's evaluation reliability in three dimensions are: RR = 0.9857, CM = 0.9911, OS = 0.9826, RM = 0.9882, PI = 0.9923, AE = 0.9656, CP = 0.9541, Mean = 0.9799. The test-retest reliability of TIMSS cognitive process attributes has reached more than 0.9, so this test has high reliability.

Results
Analysis of attribute mastery probability
The data of the four tests were analyzed. Meanwhile, in the international comparative analysis, the top 5 countries (regions) of TIMSS-2015 Grade 8 mathematics achievement were selected, which were Singapore, Korea, Taiwan, Hong Kong and Chinese mainland.

The data of the four provinces and cities in Chinese mainland were evaluated, and the results of Fig. 1 are obtained through statistical analysis.

Figure 1 illustrates the mastery of various attributes in the process of mathematical cognition among Grade 8 students across four provinces (cities) as well as the overall. On the whole, students' mastery of various attributes of cognitive process is relatively balanced, basically maintained at about 70%. In comparison, students' mastery of the attributes of *Operation and Solution* is low, only reaching more than 60%, and their mastery of the three attributes of *Recollection and Recognition*, *Calculation and Measurement* and *Generalization and Proof* is good, reaching more than 75%.

From the comparison of provinces (cities), the performance of students in Shanghai in the seven attributes of cognitive process is significantly better than those in other provinces and cities, reflecting an absolute advantage, basically reaching more than 85%, and even more than 90% in the attributes of *Recollection and Recognition* and *Generalization and Proof*. Only the mastery of *Operation and Solution* is slightly lower, only 83.2%. The data of Gansu, Guizhou and Guangdong provinces are basically consistent, which are lower than the overall level. Only Gansu's *Recollection and Recognition* and *Representation Modeling* attributes are higher than the overall level. At the same time, it can be found that students in Guangdong have the lowest probability of mastering attributes in *Recollection and Recognition*, *Calculation and Measurement* and *Generalization and Proof*, and students in Guizhou have the lowest probability of mastering attributes in *Analysis and Evaluation*. Students in Guizhou, Gansu and Guangdong have almost the same mastery of *Operation and Solution* attributes with only 61%. The two attributes of *Operation and Solution* and *Analysis and Evaluation* show low mastery probability in all provinces and cities.

Based on data analysis of mainland China, further evaluation was conducted comparing data from mainland China and the top five countries (regions). The results, presented in Fig. 2, were obtained through statistical analysis and data centralization processing.

According to Fig. 2, Chinese mainland has a higher level of cognition in seven cognitive processes. Apart from the two attributes of *Process Implementation* and *Analysis and Evaluation*, Singapore has the best performance in five other attributes, and its value reaches the maximum of six countries. Singapore has an absolute advantage in the attribute *Process Implementation*. Japan has some advantages in *Representation modeling*, which value is only inferior to that of Chinese mainland, and the other six attributes are below the average. In addition to being slightly higher than the average value in terms of *Representation Modeling* and *Process Implementation*, other attributes in South Korea are lower than the average value, and *Analysis and Evaluation* has reached the lowest value in six countries, with poor performance. In addition to RM in Taiwan is far lower than other countries. The other six cognitive process attributes are basically close to the average and are relatively balanced. Hong Kong, China is higher than the average in the attributes of *Recollection and Recognition*, *Operation and Solution*, and *Generalization and Proof*, especially in the attribute of *Recollection and Recognition*. However, the performance of *Calculation and Measurement*, *Process Implementation*, and *Analysis and Evaluation* is not good, especially
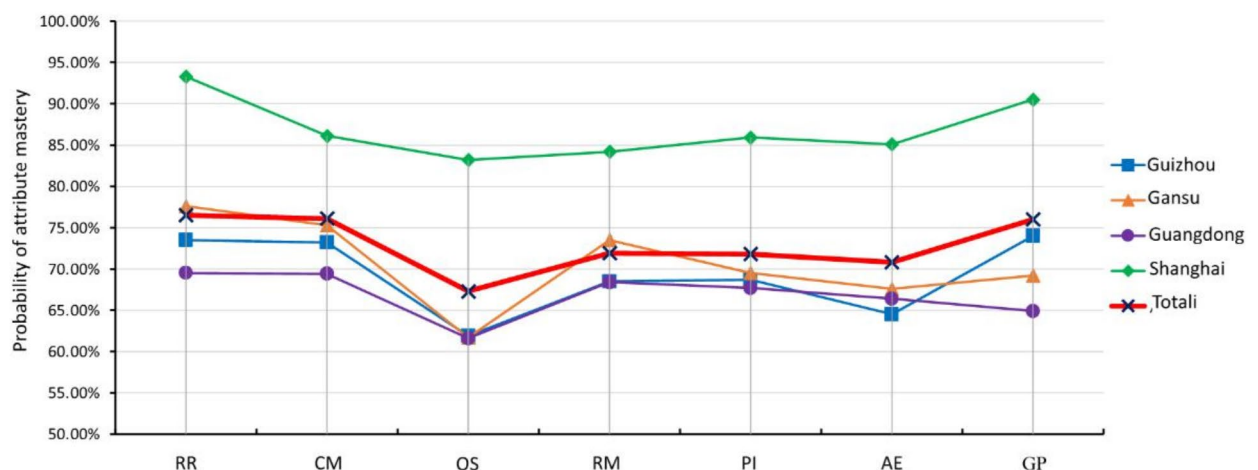
Fig. 1. Broken line chart of attribute mastery of mathematical cognition process in Grade 8 in four provinces (cities) of China. Recollection and recognition (RR), Calculation and measurement (CM), Operation and solution (OS), Representation modeling (RM), Process implementation (PI), Analysis and evaluation (AE), Generalization and Proof (GP).
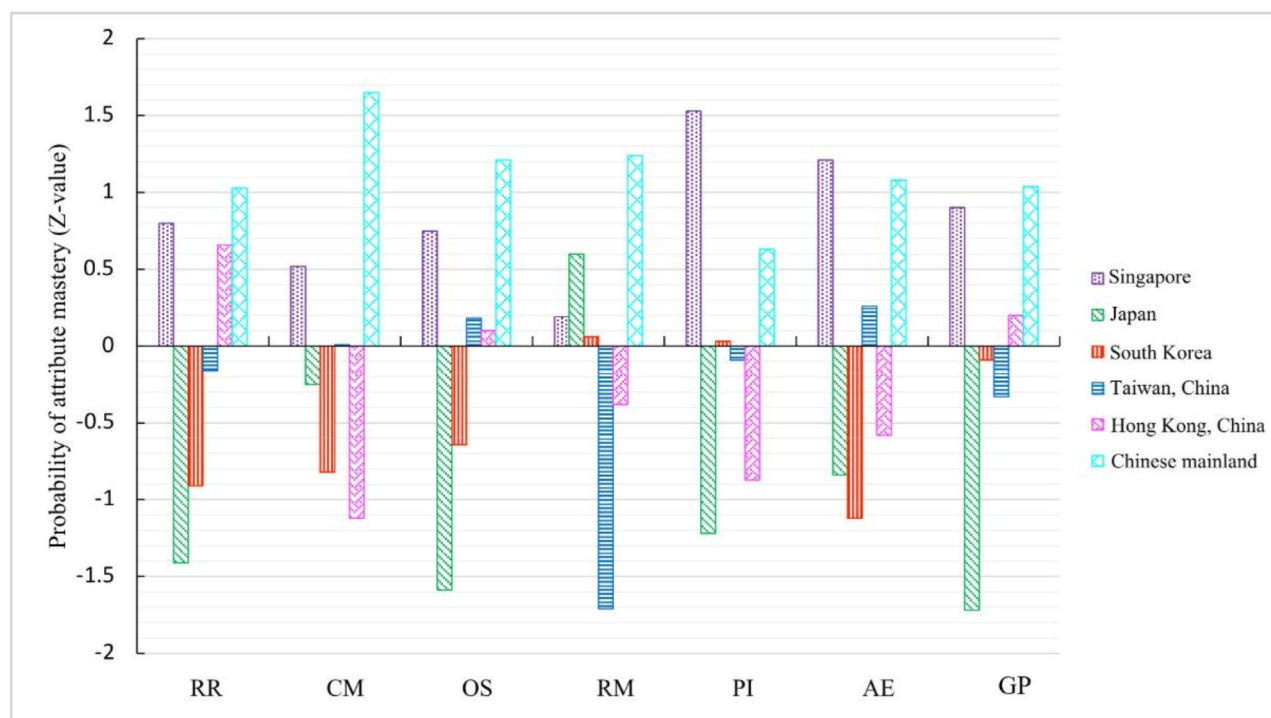
Fig. 2. Standardized distribution of eighth-grade mathematics cognitive process attribute mastery in different countries or regions. Recollection and recognition (RR), Calculation and measurement (CM), Operation and solution (OS), Representation modeling (RM), Process implementation (PI), Analysis and evaluation (AE), Generalization and Proof (GP).

when the *Calculation and Measurement* attribute reach the minimum value of six countries or regions, and the mastery attribute is the lowest.

Overall, students of Chinese mainland have absolute superiority in the control of cognitive process attributes, almost all of them are at the best level. Only two attributes of *Process Implementation* and *Analysis and Evaluation* are ranked second, which is second only to Singapore with little difference. Students of Chinese mainland have the most obvious advantages in *Calculation and Measurement*, *Operation and Solution*, and *Representation Modeling*, which are far more obvious than those from other five countries or regions.

Analysis of advanced learning

Based on the data of students' responses, the CDM evaluates the mastery of each student's different attributes, and finally makes a judgment on the mastery status of each attribute (mastery is marked as 1, non-mastery is marked as 0). Therefore, each student's mastery of different attributes forms a multidimensional vector composed of 0 or 1, which is usually called knowledge state[5]. Then, through IRT, using the three parameter IRT (3PL) and the mirt-package in R, the corresponding ability (θ) of each student under the item response theory is calculated[33]. According to the classes formed by different knowledge states, the average value of the ability values of all students in this class is calculated as the ability value of this knowledge state. The clustering results and capability values are shown in Table 4:

As shown in Table 4, the capability value corresponding to the knowledge state (00000000) is the smallest, with a value of -1.42, and the capability value corresponding to the knowledge state (11111111) is the largest, with a value of 0.56. Therefore, dividing the ability value from − 1.5 to 0.6 into five levels with every 0.42 as a level, and the learning progress chart can be obtained as shown in Fig. 3 below:

As illustrated in the learning path diagram in Fig. 3, the majority of knowledge states fall in the third level, with some distributed across the second and fourth levels, and only one knowledge state present in both the first and fifth levels. This distribution indicates that students did not progress in a linear or equidistant manner in their development of cognitive processes. Instead, there appear to be periods of significant leaps, particularly in the initial and final stages of learning. Using the same methodology for constructing advanced learning paths based on foundational knowledge and skills, the attributes associated with knowledge states at various ability levels were extracted to define advanced learning. The steps to derive advanced cognitive dimensions are summarized in Table 5 below.

The cognitive process of students is related to several elements including their interest in learning and their own learning characteristics, the logical structure of subject content and teachers' teaching. Therefore, the learning path reflected by the above data is the result of comprehensive factors, which can provide some help for guiding students' learning and teachers' teaching[35].

Knowledge state	Ability value	Variance	Knowledge state	Ability value	Variance	Knowledge state	Ability value	Variance
0000000	-1.42	0.27	1,000,101	-0.62	0.02	1,100,111	-0.23	0.02
1,000,000	-0.79	0.02	0101001	-0.66	0.00	1,101,101	-0.33	0.01
0000001	-0.83	0.03	1,001,110	-0.50	0.01	1,111,001	-0.28	0.02
1,001,000	-0.61	0.00	1,101,010	-0.40	0.01	1,101,011	-0.35	0.01
1,100,000	-0.68	0.02	1,101,100	-0.45	0.03	1,111,110	-0.25	0.00
1,000,001	-0.79	0.03	1,100,101	-0.41	0.02	1,110,111	-0.2	0.03
0100001	-0.76	0.02	1,100,011	-0.43	0.01	1,111,101	-0.21	0.02
0001001	-0.78	0.01	1,101,001	-0.44	0.01	1,111,011	-0.31	0.04
1,001,010	-0.64	0.02	0101011	-0.48	0.00	1,111,111	0.56	0.25
1,101,000	-0.61	0.00	1,101,110	-0.32	0.02			
1,100,001	-0.60	0.02	1,110,101	-0.28	0.03			

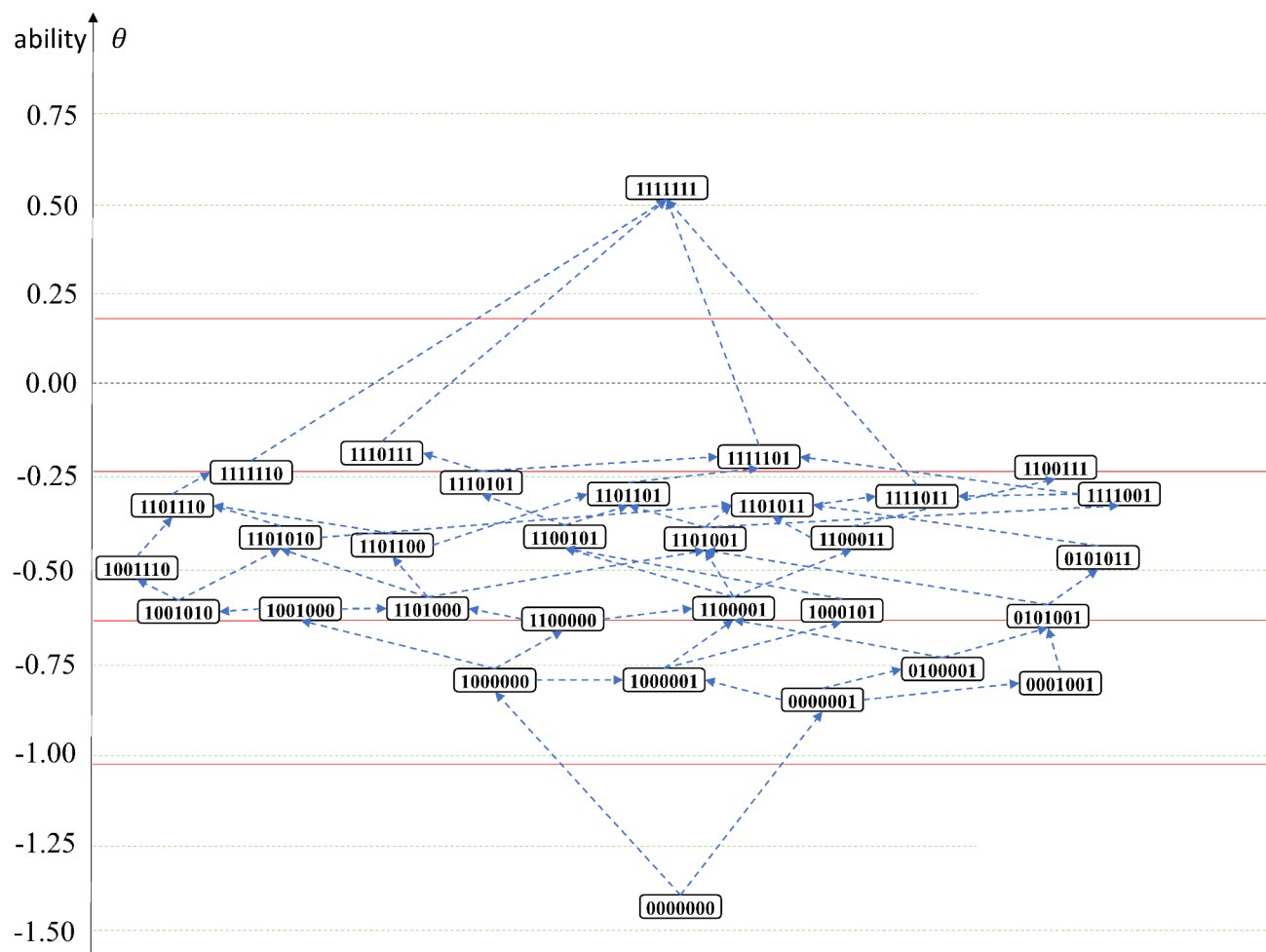**Table 4**. Cognitive process attribute values of TIMSS.

Fig. 3. Advanced learning diagram of cognitive process attributes.

Personalized analysis

In the personalized analysis, this study selects four students numbered GZBS051, GSLS252, GDZD451, SHYH026 as the research participants (e.g., Jack, Sarah, John, Lucas for each). Their common feature is that they have a common total score under the traditional test theory, but their knowledge states are (0100001), (1001010), (1100001) and (1100000) separately. Jack and Lucas only master two attributes, while Sarah and John master three attributes. There are also great differences in the types of mastering attributes, as shown in Fig. 4:

Figure 4. Comparative analysis of four students' mastery of dimension attributes of cognitive process with the same total score.

Level	Content to define	Attribute to master
1	Students have some preliminary cognition of mathematics, but have not completely mastered any cognitive attributes, which belongs to the preliminary stage of cognitive process development	No attributes are fully mastered
2	Students at this level can recall basic facts in mathematics, distinguish expressions from geometric figures, judge mathematical equivalence relationships, and classify numbers, expressions, quantities, and shapes by common attributes; At the same time, according to the expression and process of preliminary induction and generalization, with more general and more widely applicable terms to express the relationship, can provide some schemes for mathematical demonstration.	Preliminary mastery of attributes RR and GP
3	Students can perform operations on numbers and algebraic expressions, obtain useful information from resources such as graphs, tables and texts, obtain data through measurement and represent data in different forms; Be able to model practical problems by creating equations, inequalities, and geometric figures based on specific problems, and generate equivalent representations for given mathematical expressions or relations	On the basis of good command of RR, GP, preliminarily master CM and RM
4	Identify effective and appropriate actions, as well as strategies and tools to address common problems of existing solutions; Implementing strategies and operations to solve problems involving familiar mathematical concepts and processes, often involving multiple steps, reflects the process of problem solving	On the basis of the preliminary master OS and PI attributes
5	Based on a comprehensive grasp of the first four levels, identify, describe or use relationships among numbers, expressions, quantities and shapes, evaluate alternative strategies and solutions to problems, and make effective inferences based on information and evidence	Master all attributes

Table 5. Classification of learning advanced level in TIMSS cognitive process attribute. Recollection and recognition (RR), Calculation and measurement (CM), Operation and solution (OS), Representation modeling (RM), Process implementation (PI), Analysis and evaluation (AE), Generalization and Proof (GP).
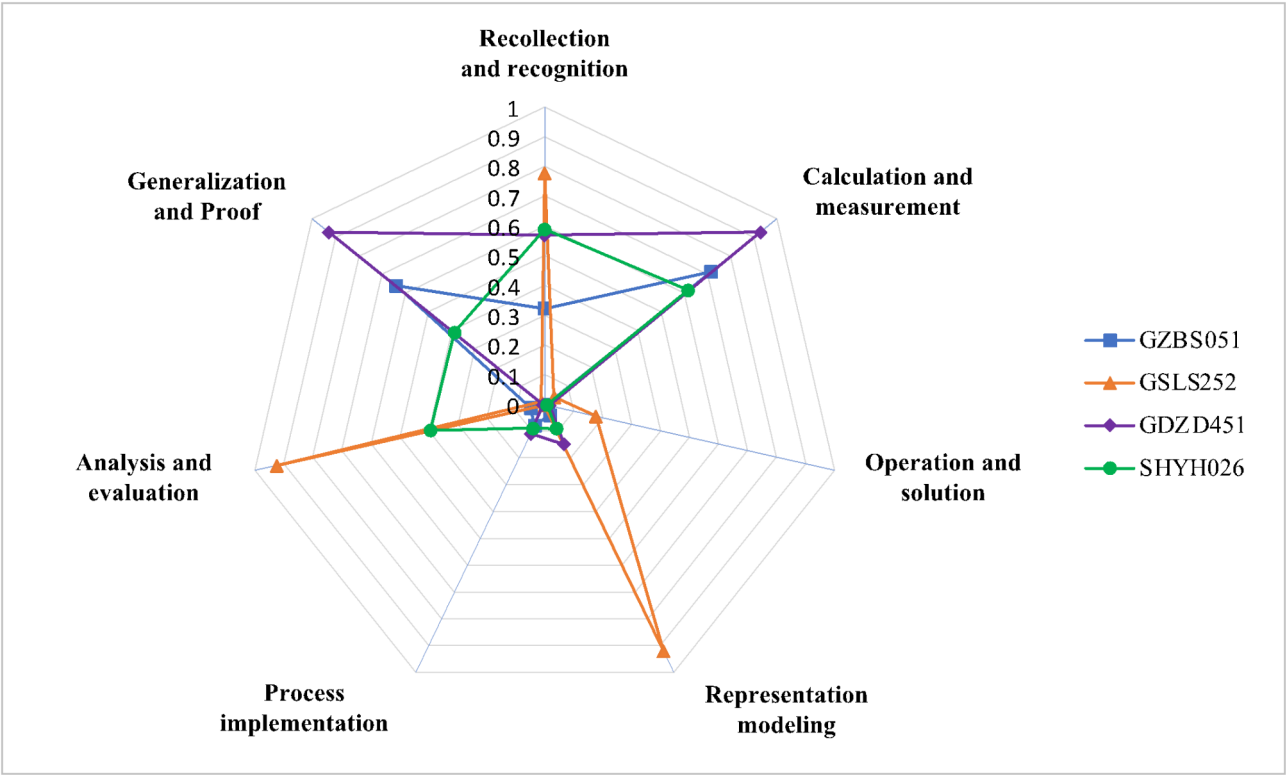
Fig. 4. Comparative analysis of four students' mastery of dimension attributes of cognitive process with the same total score. GZBS051: Jack, GSLS252: Sarah, GDZD451: John, SHYH026:Lucas.

It can be seen from Fig. 4 that the probabilities of *Calculation and Measurement* and *Generalization and Proof* attributes of Jack are about 0.7. Although the probabilities of mastering the two attributes are large, it has not been fully mastered and needs to be strengthened. In the *Recollection and Recognition* attribute knowledge states, it is not mastered, but there is also a certain probability of mastery and a certain foundation. Sarah has a high probability of mastering the attributes of *Recollection and Recognition*, *Representation Modeling,* and *Analysis and Evaluation*, which are more than 0.8, and the probability of mastering other attributes is almost close to 0. Meanwhile, John has a high probability of mastering the attributes of *Calculation and Measurement* and *Generalization and Proof*, reaching more than 0.9, but the probability of mastering the attributes of *Recollection and Recognition* is only slightly higher than 0.5, which is in the state of partial mastery. The probability of mastering other attributes is very low, which can be considered as not mastering at all. Lucas's mastery probabilities of: *Recollection and Recognition* and *Calculation and Measurement* attributes are about 0.6. In terms of knowledge state, although Lucas has mastered this attribute, it needs to be further strengthened. In terms of the attributes

of *Analysis and Evaluation* and *Generalization and Proof*, the mastery probability of Lucas is about 0.4, which means he or she has a certain foundation for this attribute, but has not reached the conditions for mastering this attribute.

Discussion

This study collected data from 4,733 Grade 8 students across four provinces (cities) in Mainland China and utilized CDA to analyze their performance across seven cognitive process attributes. Additionally, the top five countries in mathematics achievement from TIMSS-2015 for Grade 8 were selected for comparison, allowing for a detailed evaluation of Mainland China's results against these high-performing countries. The analysis thoroughly explored the data from three perspectives: attribute mastery, advanced learning progression, and personalized learning analysis. These findings serve as a pre-test for Mainland China's participation in TIMSS and provide a more standardized research framework for applying cognitive diagnostic methodologies.

First, the findings indicate that Grade 8 students in Mainland China demonstrate a significant advantage in mastering mathematical cognitive processes, particularly in *Calculation and Measurement* (CM), *Operation and Solution* (OS), and *Representation Modeling* (RM). These traditional domains are critical for solving routine problems, achieving procedural fluency, and applying mathematical concepts in structured scenarios. This reflects the effectiveness of the current curriculum and instructional practices in mainland China, which emphasize fundamental mathematical skills and applications. Such strengths suggest that students are well-prepared for standardized tests and structured problem-solving, showcasing the success of a system designed to ensure proficiency in foundational mathematics. However, while Mainland Chinese students demonstrate strong procedural fluency, the curriculum may need to integrate more inquiry-based and exploratory learning to foster higher-order thinking skills, such as critical reasoning and innovative problem-solving[36–38]. These findings underscore the need for a balanced educational approach that combines mastery of fundamental skills with opportunities for advanced mathematical reasoning and creativity.

Moreover, result suggests that students' cognitive development in mathematics is not uniform or steady but rather characterized by periods of accelerated progress, particularly during the early and advanced stages of learning. This irregular progression highlights the complexity of cognitive development[39–41], which is influenced by multiple factors such as students' intrinsic interest in learning[42,43], individual learning characteristics[43], the logical structure of the subject content[45], and the effectiveness of teaching methods[46]. These insights emphasize the importance of tailoring instruction to support diverse learning trajectories. Educators can design targeted interventions and scaffolded learning experiences to address varying rates of cognitive development, while curriculum developers can align educational content with students' natural learning paths, fostering more effective and personalized learning environments.

In the end, the personalized analysis of students' mastery probabilities, as illustrated in Fig. 4, highlights the diverse strengths and weaknesses among individual learners, underscoring the importance of personalized instruction[47–49]. For instance, Jack and Lucas require targeted support to strengthen their foundational skills in *Calculation and Measurement* and *Representation Modeling*, while Sarah and John show varying degrees of proficiency across attributes that could benefit from focused teaching strategies. By leveraging data-driven insights, educators can design adaptive learning pathways that provide feedback-driven support, enabling students to build on their strengths while addressing areas of weakness. Such approaches not only enhance individual learning outcomes but also inform broader curriculum development aimed at equitable educational opportunities.

CDA has gained prominence among educational researchers for its ability to enrich traditional evaluation methods by providing detailed diagnostic information. Nichols highlighted that CDA offers educators and decision-makers insights into students' problem-solving strategies, conceptual understanding, and mastery of domain-specific principles[50]. Previous studies applying CDA to TIMSS data have demonstrated its potential to uncover nuanced differences in mathematics performance across nations. For example, studies comparing South Korean and American students have revealed disparities in problem-solving and reasoning, as well as the impact of teacher guidance on student learning[51,52]. However, many of these studies faced limitations in attribute construction and the depth of data analysis. This study addresses these gaps by improving attribute construction and fully leveraging the diagnostic potential of the data, providing a more comprehensive understanding of students' mathematical cognitive processes.

Conclusion

With the application of cognitive diagnostic theory and the adaptation of the TIMSS test, this study developed a cognitive diagnostic tool to analyze students' mathematical cognitive processes. Through an in-depth analysis of data from Mainland China, meaningful conclusions were drawn, providing insights into students' mathematical learning. Moreover, international comparisons highlighted both the strengths and weaknesses of mathematics learning among Mainland Chinese students. Importantly, the knowledge states derived from the CDA allowed for the construction of student learning paths and progressions. While these results are theoretically sound and offer reasonable explanations for longstanding educational concerns, they remain data-driven findings. Further empirical validation, particularly through practical testing or longitudinal assessment, is necessary to confirm their alignment with real-world educational contexts.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Linn, R. L. A new era of test-based educational accountability. *Measurement: Interdisciplinary Res. Perspect.* **8**(2–3), 145–149. https://doi.org/10.1080/15366367.2010.508692 (2010).
2. Embretson, S. E. Psychometric models for learning and cognitive processes. In (eds Frederiksen, N., Mislevy, R. J. & Bejar, I. I.) *Test Theory for a new Generation of Tests* (125–150). Hillsdale, NJ: Lawrence Erlbaum Associates. (1993).
3. Leighton, J. & Gierl, M. (eds) *Cognitive Diagnostic Assessment for Education: Theory and Applications* (Cambridge University Press, 2007).
4. Chen, Y. H. *Cognitively Diagnostic Examination of Taiwanese Mathematics Achievement on TIMSS-1999* (Arizona State University, 2006).
5. Wu, X., Wu, R., Zhang, Y., Arthur, D. & Chang, H. H. Research on construction method of learning paths and learning progressions based on cognitive diagnosis assessment. *Assess. Education: Principles Policy Pract.* **28**(5–6), 657–675. https://doi.org/10.1080/0969594X.2021.1978387 (2021).
6. Ferrini-Mundy, J. & Schmidt, W. H. International comparative studies in mathematics education: Opportunities for collaboration and challenges for researchers. *J. Res. Math. Educ.* **36**(3), 164–175. https://doi.org/10.2307/30034834 (2005).
7. Chen, J. & de la Torre, J. The case of the programme for international student assessment in reading. *Psychology* **18**(5), 1967–1978. https://doi.org/10.4236/psych.2014.518200 (2014). A Procedure for Diagnostically Modeling Extant Large-Scale Assessment Data:.
8. Liu, R., Huggins-Manley, A. C. & Bulut, O. Retrofitting diagnostic classification models to responses from IRT-based assessment forms. *Educ. Psychol. Meas.* **78**(3), 357–383. https://doi.org/10.1177/0013164416685599 (2018).
9. Tatsuoka, K. K., Corter, J. & Tatsuoka, C. Patterns of diagnosed mathematical content and process skills in TIMSS-R across a sample of 20 countries. *Am. Educ. Res. J.* **41**(4), 901–926. https://doi.org/10.3102/00028312041004901 (2004).
10. Dogan, E. & Tatsuoka, K. An international comparison using a diagnostic testing model: Turkish students' profile of mathematical skills on TIMSS-R. *Educational Stud. Math.* **68**, 263–272. https://doi.org/10.1007/s10649-007-9099-8 (2008).
11. Um, E. et al. Comparing eighth-grade diagnostic test results for Korean, Czech, and American Students. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL. (2003).
12. Birenbaum, M., Tatsuoka, C. & Yamada, T. Diagnostic assessment in TIMSS-R: Between-countries and within-country comparisons of eighth graders' mathematics performance. *Stud. Educ. Eval..* **30**(2), 151–173. https://doi.org/10.1016/j.stueduc.2004.06.004 (2004).
13. Tatsuoka, K. K. Toward an integration of item-response theory and cognitive error diagnosis. In (eds Frederiksen, N., Glaser, R., Lesgold, A. & Shafto, M. G.) *Diagnostic Monitoring of Skill and Knowledge Acquisition* (453–488). Lawrence Erlbaum Associates, Inc. (1990).
14. Wu, X., Xu, T. & Zhang, Y. Research on the data analysis knowledge assessment of pre-service teachers from China based on cognitive diagnostic assessment. *Curr. Psychol.* **42**, 4885–4899. https://doi.org/10.1007/s12144-021-01836-y (2023).
15. Tatsuoka, K. K. Caution indices based on item response theory. *Psychometrika* **49**(1), 95–110. https://doi.org/10.1007/BF02294208 (1984).
16. Junker, B. W. & Sijtsma, K. Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Appl. Psychol. Meas.* **25**(3), 258–272. https://doi.org/10.1177/01466210122032064 (2001).
17. Templin, J. & Henson, R. A. *Diagnostic Measurement: Theory, Methods, and Applications* (Guilford Press, 2010).
18. Hartz, S. M. *A Bayesian Framework for the Unified Model for Assessing Cognitive Abilities: Blending Theory with Practicality* (University of Illinois at Urbana-Champaign, 2002).
19. de la Torre, J. The generalized DINA model framework. *Psychometrika* **76**(2), 179–199. https://doi.org/10.1007/s11336-011-9207-7 (2011).
20. Henson, R., Templin, J. & Willse, J. Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika* **74**(2), 191–210. https://doi.org/10.1007/s11336-008-9089-5 (2009).
21. Hagenaars, J. A. *Loglinear Models with Latent Variables* (Sage, 1993).
22. von Davier, M. Hierarchical mixtures of diagnostic models. *Psychol. Test. Assess. Model.* **52**(1), 8–28 (2010).
23. Song, L., Wang, W., Dai, H. & Ding, S. Global and item fitting indexes under cognitive diagnostic model. *Psychol. Explor.* **36**(1), 79–83 (2016).
24. Kunina-Hubenicht, O., Rupp, A. A. & Wilhelm, O. The impact of Model. Misspecification on parameter estimation. And item-fit assessment in log-linear diagnostic classification models. *J. Educ. Meas.* **49**(1), 59–81. https://doi.org/10.1111/j.1745-3984.2011.00160.x (2012).
25. Oliveri, M. E. & Von Davier, M. Investigation of model fit and score scale comparability in international. Assessments. *Psychol. Test. Assess. Model.* **53**(3), 315–333 (2011).
26. Rupp, A. A., Templin, J. & Henson, R. A. *Diagnostic Measurement: Theory, Methods, and Applications* (Guilford Press, 2010).
27. Houts, C. R. & Cai, L. *flexMIRT user's Manual Version 2.0: Flexible Multilevel Multidimensional item Analysis and test Scoring* (Vector Psychometric Group, 2013).
28. Maydeu-Olivares, A. & Joe, H. Assessing approximate fit in categorical data analysis. *Multivar. Behav. Res.* **49**(4), 305–328. https://doi.org/10.1080/00273171.2014.911075 (2014).
29. Robitzsch, A., Kiefer, T., George, A. C., Uenlue, A. & Robitzsch, M. A. Package 'CDM'. *Handbook of diagnostic classification models. New York: Springer.* (2020).
30. Hu, J., Miller, M. D., Huggins-Manley, A. C. & Chen, Y. H. Evaluation of model fit in cognitive diagnosis models. *Int. J. Test.* **16**(2), 119–141. https://doi.org/10.1080/15305058.2015.1133627 (2016).
31. Templin, J. & Beadshaw, L. Measuring the reliability of diagnostic classification model examinee. Estimates. *J. Classif.* **30**(2), 251–275. https://doi.org/10.1007/s00357-013-9129-4 (2013).
32. Tu, D., Gao, X., Liu., Cai, Y. & Wang, D. Design and implementation of cognitive diagnostic analysis system (flexCDMs). The 20th National Psychology Conference - Abstracts of Psychology and National Mental Health. (2017).
33. Wu, R. et al. A multilevel person-centered examination of students' learning anxiety and its relationship with student background and school factors. *Learn. Individual Differ..* **101**, 102253. https://doi.org/10.1016/j.lindif.2022.102253 (2023).
34. Wu, X., Zhang, Y., Wu, R., Tang, X. & Xu, T. Cognitive model construction and assessment of data analysis ability based on CDA. *Front. Psychol.*, **13**, 1009142. https://doi.org/10.3389/fpsyg.2022.1009142 (2022).
35. Nichols, P. D. A framework for developing cognitively diagnostic assessments. *Rev. Educ. Res.* **64**(4), 575–603. https://doi.org/10.3102/00346543064004575 (1994).
36. Zhang, D. & Qi, C. Reasoning and proof in eighth-grade mathematics textbooks in China. *Int. J. Educ. Res.* **98**, 77–90. https://doi.org/10.1016/j.ijer.2019.08.015 (2019).
37. Ni, Y., Li, Q., Li, X. & Zhang, Z. H. Influence of curriculum reform: An analysis of student mathematics achievement in Mainland China. *Int. J. Educ. Res.* **50**(2), 100–116. https://doi.org/10.1016/j.ijer.2011.06.005 (2011).
38. Liu, Q., Du, X., Zhao, S., Liu, J. & Cai, J. The role of memorization in students' self-reported mathematics learning: A large-scale study of Chinese eighth-grade students. *Asia Pac. Educ. Rev.* **20**(3), 361–374. https://doi.org/10.1007/s12564-019-09576-2 (2019).

39. Andrews, G. & Halford, G. S. A cognitive complexity metric applied to cognitive development. *Cogn. Psychol.* **45**(2), 153–219. https://doi.org/10.1016/S0010-0285(02)00002-6 (2002).
40. Bieri, J. & Harvey, O. J. Cognitive complexity and personality development. In *Experience Structure & Adaptability* (pp. 13–37). Springer Berlin Heidelberg. (1966). https://doi.org/10.1007/978-3-662-40230-6_3
41. Gaillard, V. & Barrouillet, P. Recent advances in relational complexity theory and its application to cognitive development. In *Cognitive Development and Working Memory* (pp. 61–82). Psychology Press. (2011). https://doi.org/10.4324/9780203845837-9
42. Renninger, K. A., Hidi, S. & Krapp, A. *The Role of Interest in Learning and Development* (L. Erlbaum Associates, 1992).
43. Krapp, A. Basic needs and the development of interest and intrinsic motivational orientations. *Learn. Instructi.*. **15**(5), 381–395. https://doi.org/10.1016/j.learninstruc.2005.07.007 (2005).
44. Bjorklund, D. F. *Children's Thinking: Developmental Function and Individual Differences* 2nd edn (Brooks/Cole Pub. Co., 1995).
45. Flavell, J. H. 1 structures, stages, and sequences in Cognitive Development. In The Concept of Development (1–28). Psychology. (2013).
46. Gallop, R. G. *The effect of student -centered and teacher -centered instruction with and without conceptual advocacy on biology students' misconceptions, achievement, attitudes toward science, and cognitive retention*. ProQuest Dissertations & Theses. (2002).
47. Walkington, C. & Bernacki, M. L. Personalization of instruction: Design dimensions and implications for cognition. *J. Experimental Educ.* **86**(1), 50–68. https://doi.org/10.1080/00220973.2017.1380590 (2018).
48. Keefe, J. W. & Jenkins, J. M. *Personalized Instruction* (Phi Delta Kappa Educational Foundation, 2005).
49. Tetzlaff, L., Schmiedek, F. & Brod, G. Developing personalized education: A dynamic framework. *Educ. Psychol. Rev.* **33**(3), 863–882. https://doi.org/10.1007/s10648-020-09570-w (2021).
50. Lee, Y. S., Choi, K. M. & Park, Y. S. Cognitive diagnosis modeling application to TIMSS: A comparison between the US and Korea via CTT, IRT, and DINA. In *Annual Meeting of the American Education Research Association, San Diego, CA*. (2009).
51. Lee, Y. S., Park, Y. S. & Taylan, D. A cognitive diagnostic modeling of attribute mastery in Massachusetts, Minnesota, and the US national sample using the TIMSS 2007. *Int. J. Test.* **11**(2), 144–177. https://doi.org/10.1080/15305058.2010.534571 (2011).
52. Im, S. & Park, H. J. A comparison of US and Korean students' mathematics skills using a cognitive diagnostic testing method: Linkage to instruction. *Educ. Res. Eval.* **16**(3), 287–301. https://doi.org/10.1080/13803611.2010.523294 (2010).

Author contributions
W.X. wrote the main manuscript text； L.N. process language of this Manuscript ; W.R. revised language and give some comments and L.H. helped revise the paper.

Funding

Declarations

Competing interests
The authors declare no competing interests.

Ethics statements
This study was approved by the Ethics Committee of East China Normal University (protocol code HR663-2022, approved on 11 November 2022).

Methods statements
In our study, all methods were carried out in accordance with relevant guidelines and regulations.
All survey (test) participants in this study obtained informed consent.

Additional information
Correspondence and requests for materials should be addressed to X.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.