Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Cross-National Cognitive Diagnostic Profiles in PISA 2015 Computer-Based Mathematics Assessment: A Joint Analysis of Attribute Mastery and Response Times

Jiwei Zhang
Northeast Normal University
Jing Lu
luj282@nenu.edu.cn

Northeast Normal University
Zhaoyuan Zhang
Yili Normal University
Guanghui Cheng
Guangzhou University

Article

Keywords: PISA 2015, Cognitive diagnosis, Mathematics content domains, Attribute mastery, Response time, Cross-national comparison, Knowledge states

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10525755/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Cross-National Cognitive Diagnostic Profiles in PISA 2015 Computer-Based Mathematics Assessment: A Joint Analysis of Attribute Mastery and Response Times

Abstract

This study examined cross-national cognitive diagnostic profiles in the PISA 2015 computer-based mathematics assessment by integrating attribute mastery and response-time information. Eleven education systems were purposively selected to reflect variation in geographic regions, educational contexts, and levels of economic development, and all students included in the analysis responded to the same set of nine test items. Specifically, this study compared students' attribute mastery probabilities across four PISA 2015 mathematics content domains, i.e., Change and Relationships, Space and Shape, Quantity, and Uncertainty and Data, examined knowledge-state distributions and cross-sectional mastery-order patterns, and evaluated response-time-informed item diagnostics, including slipping, guessing, and discrimination parameters. Using a joint cognitive diagnosis and response time modeling framework, we estimate students' mastery probabilities and knowledge-state distributions, identify cross-sectional mastery-order patterns, and evaluate item-level diagnostic characteristics. The findings indicate relatively high mastery of Change and Relationships in several education systems, substantial cross-national variation in Quantity and Uncertainty and Data, and a common cross-sectional mastery-order pattern in which Change and Relationships tends to appear earlier than Space and Shape, Quantity, and Uncertainty and Data in the distribution of knowledge states. Response-time-informed item diagnostics further reveal differences in slipping, guessing, and discrimination across items and education systems, suggesting that similar response scores may reflect different response processes and item characteristics. These results demonstrate how cognitive diagnosis and response-time modeling can extend the interpretation of PISA mathematics data beyond overall performance comparisons

and provide diagnostic information for curriculum refinement, targeted instructional support, and assessment development.

Keywords: PISA 2015; Cognitive diagnosis; Mathematics content domains; Attribute mastery; Response time; Cross-national comparison; Knowledge states

1. Introduction

Mathematical knowledge and critical thinking skills are essential for enhancing a country's competitiveness in an increasingly globalized economy (Hanushek & Woessmann, 2011). As international educational assessments have become important sources of evidence for evaluating student achievement, many education systems have paid increasing attention to students' performance in large-scale assessments and have explored ways to improve mathematics instruction in response to global educational challenges (Fischman, Topper, Silova, Goebel, & Holloway, 2019; Nortvedt, 2018). In this regard, international assessments such as the Programme for International Student Assessment (PISA) provide valuable comparative information and offer a standardized reference framework for curriculum design, assessment interpretation, and evidence-informed educational decision-making (Li, Xue, & Guo, 2025; OECD, 2016; Santos & Centeno, 2023; Zhan et al., 2018b). However, most interpretations of PISA mathematics results still rely primarily on overall achievement scores or proficiency-level summaries. Such summaries are useful for cross-national comparison, but they provide limited information about students' mastery of specific mathematics content domains. To address this limitation, the present study uses PISA 2015 computer-based mathematics assessment data to examine students' cognitive diagnostic profiles across eleven education systems. Specifically, it investigates students' mastery probabilities and knowledge-state distributions across four PISA mathematics content domains:

Change and Relationships, Space and Shape, Quantity, and Uncertainty and Data. The findings are expected to provide more fine-grained diagnostic information for curriculum refinement, targeted instructional support, and assessment development.

The Organisation for Economic Co-operation and Development (OECD) conducts the Programme for International Student Assessment (PISA) every three years. PISA is an international large-scale assessment designed to evaluate the knowledge and skills of 15-year-old students in participating countries and economies. Its overall goal is not merely to assess students' memorization of school knowledge, but to examine how well they can apply what they have learned to real-life problems and unfamiliar contexts. PISA therefore provides systematic data for comparing student performance and educational outcomes across participating education systems. These data offer important empirical evidence for policymakers, researchers, school leaders, and teachers who are concerned with educational quality and improvement. In this study, we analyze data from the PISA 2015 computer-based mathematics assessment (OECD, 2016). Item responses coded as "Not Reached" (original code 6) or "No Response" (original code 9) were excluded from the analytic dataset. Thus, the present analysis focuses on students with complete responses to the selected mathematics items. The final analytic sample consists of 11,531 students from eleven education systems. These education systems include B-S-J-G China, Japan, and Singapore in Asia; the United Kingdom, Russia, and Finland in Europe; the United States and Canada in North America; Brazil and Colombia in South America; and Australia in Oceania. For all students included in the analytic sample, both dichotomous item responses and continuous response times are available. According to the PISA 2015 mathematics assessment framework (OECD, 2016), the selected nine items represent four mathematics content domains: Change and Relationships, Space and Shape, Quantity, and Uncertainty and Data. In the present cognitive diagnosis framework,

these four PISA mathematics content domains are operationalized as diagnostic attributes, allowing students' mastery probabilities and knowledge states to be estimated for each domain.

PISA data are not only useful for estimating overall performance or latent ability, but also have the potential to support more detailed diagnostic interpretation. In particular, cognitive diagnostic assessment provides a framework for moving beyond total scores and identifying the specific knowledge components or skills that students are likely to have mastered or not mastered. Because students' internal cognitive processes cannot be directly observed from item responses alone, cognitive diagnosis models provide a model-based approach for estimating latent mastery profiles from observed response patterns. In this study, students' cognitive diagnostic profiles refer to their attribute mastery probabilities, knowledge-state classifications, and related item-level diagnostic information estimated from the cognitive diagnosis model across the four PISA mathematics content domains. This perspective is consistent with the broader goal of cognitive diagnostic assessment, which aims to provide more detailed diagnostic information than traditional score-based reporting (Tatsuoka, 1984, 2002; Maris, 1999; Junker & Sijtsma, 2001; de la Torre & Douglas, 2004; Templin & Henson, 2006; de la Torre, 2009, 2011; von Davier, 2014). A further advantage of PISA computer-based assessment is that response times can be recorded automatically for each item (OECD, 2013, 2016). Response times provide supplementary information about students' response behavior and item characteristics. Previous research has shown that response time information can be used to study response speed, test-taking motivation, response strategies, rapid guessing, test speededness, and other aspects of item response behavior (Schnipke & Scrams, 1997; van der Linden, 2007; Wise & Kong, 2005; Wang & Xu, 2015; Lu, Wang, Zhang, & Tao, 2020). In the present study, response time is incorporated as supplementary diagnostic information in a joint cognitive diagnosis and response-time modeling framework,

helping to interpret item-level diagnostic characteristics such as slipping, guessing, and discrimination.

To provide more refined diagnostic information by incorporating response-time data, this study applies a joint cognitive diagnosis and response-time modeling framework to the PISA 2015 computer-based mathematics assessment data from eleven education systems. Within this framework, the four PISA mathematics content domains—Change and Relationships, Space and Shape, Quantity, and Uncertainty and Data—are operationalized as diagnostic attributes, and response-time information is incorporated as supplementary evidence for interpreting item-level diagnostic characteristics. Accordingly, this study addresses the following three research questions:

Q1: What are the differences in students' attribute mastery probabilities across the four PISA 2015 mathematics content domains in different education systems?

Q2: What knowledge-state distributions and cross-sectional mastery-order patterns can be identified using the cognitive diagnosis model?

Q3: How do response-time-informed item diagnostics, including slipping, guessing, and discrimination parameters, help explain diagnostic differences across education systems?

Because PISA 2015 is a cross-sectional assessment, the knowledge-state distributions and mastery-order patterns examined in this study should not be interpreted as longitudinal learning trajectories or as evidence that the same students moved from one knowledge state to another over time. Through these analyses, the study aims to extend the interpretation of PISA mathematics data beyond overall performance comparisons and to generate diagnostic evidence that may inform curriculum refinement, targeted instructional support, and assessment development.

2. Important Concepts

2.1. PISA Mathematics Content Domains and Diagnostic Attributes

In the official PISA 2015 mathematics assessment framework, Change and Relationships, Space and Shape, Quantity, and Uncertainty and Data are defined as mathematics content domains rather than as cognitive attributes or mathematical attributes. These domains describe broad areas of mathematical literacy that students are expected to apply when solving problems in real-life or unfamiliar contexts. In the present study, these four PISA mathematics content domains are operationalized as diagnostic attributes within the cognitive diagnosis model. Thus, the term diagnostic attributes refers to the binary latent mastery variables used in the model to represent whether a student has mastered each corresponding mathematics content domain. In cognitive diagnostic assessment, diagnostic attributes are typically defined as specific knowledge components, skills, or content-domain mastery variables required to solve test items. Each diagnostic attribute is usually represented as a binary latent variable, with 1 indicating mastery and 0 indicating non-mastery. The relationship between test items and diagnostic attributes is specified by the Q-matrix, which indicates which diagnostic attributes are required by each item. In this study, the Q-matrix links the selected PISA mathematics items to the four PISA mathematics content domains. This modeling framework makes it possible to estimate students' mastery status for each domain and to construct model-based cognitive diagnostic profiles. These profiles provide information about students' likely strengths and weaknesses across mathematics content domains, rather than directly observed internal cognitive processes.

2.2. Attribute Mastery Probability

Attribute mastery probability refers to the posterior probability that a student has mastered a given diagnostic attribute, conditional on the observed item responses and the fitted cognitive

diagnosis model. In the present study, each diagnostic attribute corresponds to one of the four PISA mathematics content domains: Change and Relationships, Space and Shape, Quantity, and Uncertainty and Data. Therefore, the attribute mastery probability provides a model-based estimate of the extent to which a student is likely to have mastered each content domain. A higher attribute mastery probability indicates stronger evidence that the student has mastered the corresponding domain, whereas a lower probability indicates weaker evidence of mastery. Compared with total-score reporting, attribute mastery probabilities provide more fine-grained diagnostic information because they describe students' likely mastery status across multiple mathematics content domains. In this study, these probabilities are used to compare cognitive diagnostic profiles across education systems and to identify content domains in which students may require additional instructional support.

2.3 Item Level Diagnostic Parameters

Item level diagnostic parameters describe how each item functions within the cognitive diagnosis model. In this study, particular attention is given to slipping, guessing, and discrimination parameters. The slipping parameter reflects the probability that a student who has mastered the required diagnostic attribute or attributes still answers the item incorrectly. A high slipping probability may indicate that the item is difficult, cognitively demanding, linguistically complex, or vulnerable to misunderstanding. The guessing parameter reflects the probability that a student who has not mastered the required diagnostic attribute or attributes still answers the item correctly. A high guessing probability may suggest that the item provides limited diagnostic information for distinguishing mastery from non-mastery. The discrimination parameter reflects how effectively an item differentiates students with different levels of latent mastery or ability. Items with higher discrimination provide stronger information for distinguishing students with

different diagnostic profiles, whereas items with lower discrimination contribute less effectively to diagnostic classification. In the present joint cognitive diagnosis and response time modeling framework, response time information is incorporated as supplementary diagnostic information to help interpret item level response processes. Thus, item level diagnostic parameters, including slipping, guessing, and discrimination, are used to evaluate the diagnostic quality of the selected PISA mathematics items across education systems.

2.4. Knowledge States

A knowledge state refers to the combination of mastered and non-mastered diagnostic attributes for a student at the time of assessment. In a cognitive diagnosis model with four binary diagnostic attributes, each student can be classified into one of several possible knowledge states. For example, a knowledge state such as 1000 indicates that the student is estimated to have mastered the first diagnostic attribute but not the remaining three attributes, whereas a knowledge state such as 1111 indicates estimated mastery of all four diagnostic attributes. In the present study, knowledge states are defined with respect to the four PISA mathematics content domains that are operationalized as diagnostic attributes. The distribution of knowledge states within each education system provides information about how different combinations of content-domain mastery are represented among students. However, because PISA 2015 is a cross-sectional assessment, these knowledge states should be interpreted as model-based diagnostic classifications at a single assessment point. They should not be interpreted as direct evidence of students' longitudinal cognitive development or as observed transitions from one state to another over time.

2.5. Knowledge-State Patterns and Cross-Sectional Mastery-Order Interpretation

Learning trajectories, in the strict sense, refer to developmental pathways through which students' understanding progresses over time. Prior research has used the concept of learning

trajectories to describe hypothesized or empirically observed sequences of learning goals, learning activities, and changes in students' thinking during instruction (Gravemeijer, Bowers, & Stephan, 2003; Merritt, Krajcik, & Shwartz, 2008; Songer, Kelcey, & Gotwals, 2009; Lestari & Marsigit, 2020). Such trajectories generally require longitudinal evidence, repeated measurements, instructional intervention data, or other forms of developmental information that can support claims about change over time.

The present study does not examine learning trajectories in this strict longitudinal sense. PISA 2015 is a cross-sectional assessment and does not follow the same students across multiple time points. Therefore, it cannot provide direct evidence that individual students move from one knowledge state to another over time. To avoid overinterpretation, the present study uses the term cross-sectional mastery-order patterns rather than learning trajectories. These patterns are inferred from the distribution of knowledge states at a single assessment point. For example, if certain knowledge states such as 0000, 1000, 1100, 1110, and 1111 occur frequently and can be arranged from lower to higher levels of content-domain mastery, this arrangement may be interpreted as a cross-sectional mastery-order pattern. Such patterns have diagnostic value because they describe how mastery of different PISA mathematics content domains tends to be combined within an education system. They may help identify which domains appear more frequently in lower-mastery or higher-mastery profiles and may generate hypotheses about possible instructional sequencing or content-domain dependencies. However, these patterns should not be interpreted as longitudinal learning trajectories or as evidence of actual developmental progressions. Future studies using longitudinal assessment data, repeated-measure designs, or instructional intervention data would be needed to examine learning trajectories in the strict sense.

3. Statistical Models

3.1. Cognitive Diagnostic Models

Since the development of cognitive diagnostic assessment, a large number of cognitive diagnosis models have been proposed to provide fine-grained information about students' mastery of specific knowledge components or skills. These models differ in their assumptions about how diagnostic attributes contribute to item responses, as well as in their parameterization and level of complexity. Cognitive diagnosis models are often classified into compensatory, non-compensatory, and general models. Compensatory models assume that mastery of some required diagnostic attributes may compensate for non-mastery of others. A representative compensatory model is the deterministic inputs, noisy "or" gate model (DINO; Templin & Henson, 2006). In contrast, non-compensatory models assume that a correct response requires mastery of all diagnostic attributes required by an item. Representative non-compensatory models include the deterministic inputs, noisy "and" gate model (DINA; Haertel, 1989; Junker & Sijtsma, 2001) and the reduced reparameterized unified model (rRUM; Hartz, 2002). General models, such as the generalized DINA model (G-DINA; de la Torre, 2011), the log-linear cognitive diagnosis model (LCDM; Henson, Templin, & Willse, 2009), and the general diagnostic model (GDM; von Davier, 2008), allow more flexible main and interaction effects among diagnostic attributes. Under certain parameter restrictions, several specific cognitive diagnosis models can be represented as special cases of more general frameworks.

The present study uses a DINA-type cognitive diagnosis model because the primary objective is not to compare competing cognitive diagnosis models, but to estimate students' mastery probabilities and knowledge states across the four PISA mathematics content domains. In the official PISA 2015 mathematics framework, Change and Relationships, Space and Shape, Quantity, and Uncertainty and Data are defined as mathematics content domains. In this study,

these domains are operationalized as diagnostic attributes in the cognitive diagnosis model. Because each selected item is linked to one primary content domain, a parsimonious DINA-type specification provides a transparent way to estimate whether students are likely to have mastered the diagnostic attribute required by each item. More complex models, such as G-DINA, LCDM, and rRUM, are valuable when multiple diagnostic attributes jointly contribute to item responses. However, in the present application, where the selected items are each associated with a single primary diagnostic attribute, such models would introduce additional parameters without necessarily providing clearer diagnostic interpretation. Therefore, the DINA-type model is used as a parsimonious and interpretable model for estimating attribute mastery probabilities, knowledge states, and item-level diagnostic parameters.

3.2. Response Time Models

With the development of computer-based assessment, response times can be collected automatically and accurately for each item. Response-time information has long been used in psychology, educational measurement, and psychometrics to study response behavior and item characteristics. In educational assessment, response times provide supplementary information about how students interact with test items. Previous studies have used response-time information to examine response speed, test-taking motivation, response strategies, rapid guessing, test speededness, and item functioning (Schnipke & Scrams, 1997; van der Linden, 2007; Wise & Kong, 2005; Wang & Xu, 2015; Bolsinova & Tijmstra, 2018; De Boeck & Jeon, 2019). In the present study, response time is not treated as an independent cultural variable. Rather, it is incorporated as supplementary diagnostic information to help interpret item level response processes and diagnostic item characteristics.

Raw response times are non-negative and usually positively skewed. Therefore, various distributions and transformations have been used to model response-time data, including exponential distributions (Scheiblechner, 1979), gamma distributions (Maris, 1993), Weibull distributions (Rouder et al., 2003), and log transformations (van der Linden, 2007). Semi-parametric models, such as proportional hazards models and linear transformation models, have also been proposed for response-time analysis (Ranger & Ortner, 2012; Wang, Fan, Chang, & Douglas, 2013; Wang, Chang, & Douglas, 2013). Among these approaches, the log-normal response-time model is widely used in psychometric research because the logarithmic transformation often makes positively skewed response-time distributions more symmetric and suitable for normal modeling.

In this study, we adopt the log-normal response time model to analyze item level response times in the PISA 2015 computer-based mathematics assessment. This model is appropriate because it accounts for the non-negativity and right-skewness of response-time data and provides interpretable person- and item-level parameters. Specifically, the model estimates item time intensity and person speed, which help describe how much time an item tends to require and how quickly a student tends to work. Prior studies have demonstrated the usefulness of log-normal response-time models in educational and psychometric applications (Fox et al., 2007; Klein Entink et al., 2009; Wang & Xu, 2015; Qian, 2016; Bolsinova et al., 2017; Zhan et al., 2018). Therefore, the log-normal response time model is used here as a suitable and interpretable framework for incorporating response-time information into the cognitive diagnostic analysis.

3.3. Joint Model for Cognitive Diagnosis and Response Time

To jointly analyze item responses and response times, this study combines a DINA-type cognitive diagnosis model with a log-normal response-time model. The cognitive diagnosis

component estimates students' mastery probabilities and knowledge states across the four PISA mathematics content domains, whereas the response time component provides supplementary information about students' response speed and item time intensity. The joint framework allows item responses and response times to be analyzed within a coherent model, thereby supporting response-time-informed item diagnostics, including slipping, guessing, and discrimination parameters.

3.3.1. Cognitive Diagnosis Measurement Model: Reparameterized DINA Model

Let Y_{ij} denote the observed dichotomous response of examinee *i* to item *j*, where *i*=1,…,*N* and *j*=1,…,*J*. Let $Y_{ij} = 1$ indicate a correct response and $Y_{ij} = 0$ indicate an incorrect response.

In the DINA model, the ideal response indicator is defined as

$$\eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}},$$

where α_{ik} is a binary latent mastery variable indicating whether examinnee *i* has mastered diagnostic attribute *k*, and q_{jk} is an element of the Q-matrix indicating whether item *j* requires diagnostic attribute *k*. Specifically, $q_{jk} = 1$ means that attribute *k* is required by item *j*, whereas $q_{jk} = 0$ means that it is not required. In this study, the diagnostic attributes correspond to the four PISA mathematics content domains operationalized in the cognitive diagnosis model. Given the ideal response indicator η_{ij}, the probability of a correct response is

$$p(Y_{ij} = 1 \mid \boldsymbol{\alpha}_i) = g_j + (1 - s_j - g_j)\eta_{ij},$$

where s_j is the slipping parameter and g_j is the guessing parameter. The slipping parameter s_j represents the probability that a student who has mastered the required diagnostic attribute still answers item *j* incorrectly, whereas the guessing parameter g_j represents the probability that a student who has not mastered the required diagnostic attribute still answers item *j* correctly. The

item discrimination index can be defined as $1 - s_j - g_j = \text{IDI}_j$ (de la Torre, 2008). Based on the traditional DINA model, we reparameterize s_j and g_j from the probability scale to the logit scale (DeCarlo, 2011). That is,

$$\zeta_j = \text{logit}(g_j), \tag{2}$$

$$\beta_j = \text{logit}(1 - s_j) - \text{logit}(g_j), \tag{3}$$

where $\text{logit}(x) = \log(x/(1-x))$. Therefore, the reparameterized DINA model (DeCarlo, 2011; Qiao & Jiao, 2021) can be written as

$$\text{logit}\left[p\left(Y_{ij} = 1 \mid \boldsymbol{\alpha}_i, \varsigma_j, \beta_j\right)\right] = \varsigma_j + \beta_j \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}}, \tag{4}$$

where ς_j and β_j are the item intercept and interaction parameters, respectively.

3.3.2. Higher-Order Latent Structural Model

To model the relationship among the diagnostic attributes, a higher-order latent structural model is used. Let θ_i denote the general latent ability of examinee *i*. The probability that examinee *i* has mastered diagnostic attribute *k* is modeled as

$$p(\alpha_{ik} = 1 \mid \theta_i, \xi_k, \varpi_k) = \frac{\exp(\xi_k \theta_i - \varpi_k)}{1 + \exp(\xi_k \theta_i - \varpi_k)}, \tag{5}$$

where ξ_k is the slope parameter for diagnostic attribute *k*, and ϖ_k is the corresponding intercept parameter. This higher-order structure links examinees' general latent ability to their mastery probabilities across the four diagnostic attributes. In the present study, these diagnostic attributes represent the four PISA mathematics content domains within the cognitive diagnosis model.

3.3.3. Response Time Model

Let T_{ij} denote the response time of examinee *i* on item *j*. Because response times are non-negative and usually positively skewed, the logarithm of response time is modeled using a normal distribution. Following the log-normal response time model, we specify

$$\log(T_{ij}) = \lambda_j - \tau_i + e_{ij},\ e_{ij} \sim N(0, \sigma_j^2), \tag{6}$$

where λ_j is the time intensity parameter of item j. A higher value of λ_j indicates that the item tends to require more time. The parameter τ_i represents the speed of examinee i. A higher value of τ_i indicates faster response speed and therefore a shorter expected response time. The residual term e_{ij} is assumed to follow a normal distribution with mean zero and item specific variance σ_j^2, allowing the variability of log response times to differ across items.

3.3.4. Hierarchical Structural Model

To capture the across-person relationship between speed and accuracy (van der Linden, 2006), the ability and speed parameters are assumed to follow a bivariate normal distribution in the population, that is,

$$\boldsymbol{\xi}_i = (\theta_i, \tau_i)' \sim N(\boldsymbol{\mu}_P, \boldsymbol{\Sigma}_P), \tag{7}$$

with mean vector

$$\boldsymbol{\mu}_P = (\mu_\theta, \mu_\tau)', \tag{8}$$

and covariance matrix

$$\boldsymbol{\Sigma}_P = \begin{pmatrix} \sigma_\theta^2 & \sigma_{\theta\tau} \\ \sigma_{\tau\theta} & \sigma_\tau^2 \end{pmatrix}. \tag{9}$$

The covariance parameter $\sigma_{\theta\tau}$ describes the population level association between latent ability and response speed. A positive covariance indicates that students with higher latent ability tend to respond faster, whereas a negative covariance indicates that examinees with higher latent ability tend to respond more slowly. This structure allows the joint model to examine the relationship between accuracy related latent ability and response time behavior without treating response time as an independent cultural variable.

4. Bayesian Statistical Inference

4.1. Model Identification

Several identification constraints were imposed to ensure stable estimation of the joint cognitive diagnosis and response-time model. First, to eliminate the location indeterminacy between the higher-order latent ability parameter θ_i and the diagnostic attribute intercept parameter ϖ_k, the population mean of the latent ability parameter was fixed at zero, that is, $\mu_\theta = 0$. Similarly, to remove the location indeterminacy between the person speed parameter τ_i and the item time intensity parameter λ_j, the population mean of the speed parameter was fixed at zero, that is, $\mu_\tau = 0$. Second, the variance of the latent ability parameter was fixed at one, $\sigma_\theta^2 = 1$, to set the scale of the higher-order latent ability dimension and to avoid scale indeterminacy between θ_i and the diagnostic attribute slope parameter ξ_k(Lord & Novick, 1968). The diagnostic attribute slope parameter was constrained to be positive, $\xi_k > 0$, so that a higher level of latent ability corresponds to a higher probability of mastering the corresponding diagnostic attribute. In addition, to ensure positive item discrimination in the reparameterized DINA model, the item discrimination index was required to satisfy $\mathrm{IDI}_j > 0\big(1 - s_j > g_j\big)$, which is equivalent to $1 - s_j > g_j$. Under the logit-scale reparameterization, this condition corresponds to the restriction β_j>0. Therefore, the prior distribution for β_j was specified so that sampled values were restricted to the positive domain (Zhan, Jiao, & Liao, 2018a). These constraints identify the ability, speed, diagnostic attribute, and item level diagnostic parameters and allow the posterior estimates to be interpreted consistently across education systems.

4.2. Prior Distributions and Parameter Estimation

The priors and hyper-priors for the item parameters are assigned the same values as in Zhan et al. (2018a). Model parameters were estimated in a full Bayesian framework using JAGS (Plummer, 2015), called from R through the R2jags package (Su & Yajima, 2015). JAGS adopts the slice-with-Gibbs sampler to implement Markov chain Monte Carlo (MCMC) sampling. Posterior estimation was conducted using Markov chain Monte Carlo (MCMC) sampling. Each chain was run for 10,000 iterations, with the first 5,000 iterations discarded as burn-in. Convergence was assessed using the potential scale reduction factor (PSRF; Brooks & Gelman, 1998) for all monitored parameters.

5. Materials and Research Analysis

This study uses data from the PISA 2015 computer-based mathematics assessment (OECD, 2016). The analysis is designed to examine students' cognitive diagnostic profiles across eleven education systems, rather than to make causal claims about cultural or economic differences. The eleven education systems were purposively selected to reflect variation in geographic regions, educational contexts, and levels of economic development (Nyimbili & Nyimbili, 2024; Patton, 2002). This selection allows the study to describe cross-national variation in students' mastery of PISA mathematics content domains while maintaining a common assessment basis. Because PISA 2015 is a cross-sectional assessment, the present analysis does not examine longitudinal learning trajectories or developmental changes over time. Instead, it focuses on attribute mastery probabilities, knowledge-state distributions, cross-sectional mastery-order patterns, and response-time-informed item diagnostics.

In PISA 2015, students from different education systems may have responded to different test forms or item sets. To ensure comparability across education systems, the present study focuses only on students who responded to the same set of nine computer-based mathematics items.

These items are CM033Q01, CM474Q01, CM155Q01, CM155Q04, CM411Q01, CM411Q02, CM803Q01, CM442Q02, and CM034Q01. These items were selected because their links to the PISA mathematics content domains could be clearly identified from the publicly available PISA documentation, allowing a corresponding Q-matrix to be constructed. By restricting the analysis to this common item set, the study ensures that comparisons of attribute mastery probabilities, knowledge-state distributions, and item level diagnostic parameters are based on the same assessment information across education systems. The final analytic sample consists of 11,531 students from eleven education systems. These education systems include B-S-J-G China (CHI, 1,067 students), Singapore (SGP, 687), and Japan (JPN, 731) in Asia; the United States (USA, 567) and Canada (CAN, 2,021) in North America; the United Kingdom (GBR, 1,396), Finland (FIN, 603), and Russia (RUS, 571) in Europe; Australia (AUS, 1,422) in Oceania; and Brazil (BRA, 1,365) and Colombia (COL, 1,101) in South America. The term education systems is used because the Chinese sample in PISA 2015 refers specifically to B-S-J-G China, namely Beijing, Shanghai, Jiangsu, and Guangdong, rather than to a nationally representative sample of mainland China.

For each student in the analytic sample, dichotomous item responses and response times were available for the selected items. Item responses coded as “Not Reached” and “No Response” were excluded from the analytic dataset, and only students with complete responses to the nine selected items were retained. For student i and item j, a correct response was coded as Y_{ij}=1, whereas an incorrect response was coded as Y_{ij}=0. These dichotomous response data were used to estimate students’ attribute mastery probabilities and knowledge states within the cognitive diagnosis model. Response time was recorded automatically by the computer-based testing system and represents the time elapsed between item presentation and response submission. In the present study, response time is treated as supplementary diagnostic information for interpreting item-level

response processes and item characteristics, rather than as a direct measure of students' internal cognitive processes or as an independent cultural variable. The item response and response-time data are available from the public PISA database (https://www.oecd.org/pisa/data/.). The variable system in this study is based on the four PISA 2015 mathematics content domains: Change and Relationships, Space and Shape, Quantity, and Uncertainty and Data. In the official PISA framework, these domains represent broad areas of mathematical literacy. In the present cognitive diagnosis model, they are operationalized as diagnostic attributes. This distinction is important because the study does not claim that PISA officially defines these domains as cognitive attributes. Rather, the four content domains are used as model-based diagnostic attributes to estimate students' mastery probabilities and knowledge states.

The four mathematics content domains can be interpreted as follows. "Change and Relationships" concerns algebraic, functional, and relational reasoning, including the understanding of variable dependencies and dynamic relationships. "Space and Shape" concerns geometric representation, spatial visualization, and visual reasoning. Quantity concerns numerical reasoning, measurement, comparison, and quantitative modeling. "Uncertainty and Data" concerns probabilistic thinking, data interpretation, variability, and reasoning under uncertainty. These four content domains are listed in Table 1 and are used as diagnostic attributes in the cognitive diagnosis model. The Q-matrix in Table 2 specifies the relationship between the selected items and the four diagnostic attributes. A value of 1 in the Q-matrix indicates that the corresponding item is linked to the diagnostic attribute, whereas a value of 0 indicates that it is not. Although broader mathematical processes such as reasoning and problem solving are important components of mathematical literacy, the publicly available item-level documentation does not provide sufficient information to construct a defensible Q-matrix for these additional process

dimensions in the present item set. Therefore, the current analysis focuses on the four PISA mathematics content domains to ensure conceptual clarity and modeling consistency.

5.1. Descriptive Statistical Analysis

A preliminary descriptive analysis was conducted for the item response and response-time data. Table 3 presents the number, percentage, and cumulative percentage of correct responses for students across the eleven education systems. These descriptive results provide an initial overview of the correct-response distributions before the cognitive diagnosis and response-time models are fitted.

The distributions of total correct responses varied across education systems. Among the Asian education systems, relatively large proportions of students in B-S-J-G China, Singapore, and Japan answered fewer than five or six of the nine items correctly. In North America, 57.20% of students in Canada answered fewer than five items correctly, whereas a relatively higher proportion of students in the United States answered fewer than four items correctly. In Europe, 44.90% of students in the United Kingdom answered fewer than four items correctly, and 47.90% of students in Finland answered fewer than five items correctly. The two South American education systems showed lower correct-response counts, with 54.23% of students in Brazil and 55.10% of students in Colombia answering fewer than four and two items correctly, respectively. In Oceania, 58.08% of students in Australia answered fewer than five items correctly. The highest percentages of students answering all nine items correctly were observed in Japan (9.20%), Singapore (9.00%), and Finland (6.00%), whereas the lowest percentages were observed in Colombia (0.50%) and Russia (2.00%). These descriptive differences suggest that the selected education systems differ in their overall correct-response distributions, which provides a useful background for interpreting the subsequent diagnostic results.

The items with the highest average response times were CM442Q02, CM411Q01, and CM034Q01, with average response times ranging from 1.616 to 3.332 minutes across education systems. Japan, Singapore, and Australia showed relatively shorter average response times for these items. For item CM474Q01, the average response times were relatively shorter in Australia, Finland, the United Kingdom, B-S-J-G China, Japan, and Singapore. For item CM033Q01, Canada, the United States, and Russia showed relatively shorter average response times. For item CM411Q02, Brazil and Colombia showed relatively shorter average response times. These descriptive response-time patterns are not interpreted as direct evidence of cultural differences in test-taking behavior. Rather, they provide preliminary information about item-level response behavior and support the later response-time-informed item diagnostic analysis.

5.2. Attribute Mastery Probability

In this section, we examine students' estimated attribute mastery probabilities across the four PISA mathematics content domains, which were operationalized as diagnostic attributes in the cognitive diagnosis model. Table 4 and Figure 1 present the average posterior mastery probabilities for Change and Relationships, Space and Shape, Quantity, and Uncertainty and Data across the eleven education systems.

The estimated mastery probabilities varied across both education systems and mathematics content domains. Singapore showed the highest average mastery probabilities for all four domains: Change and Relationships (0.826), Space and Shape (0.679), Quantity (0.701), and Uncertainty and Data (0.625). B-S-J-G China showed a slightly higher estimated mastery probability than Japan in Change and Relationships, whereas Japan showed higher estimated mastery probabilities in the other three domains. Brazil showed the lowest estimated mastery probabilities across the four domains, with values of 0.353, 0.264, 0.230, and 0.085, respectively. Colombia showed

slightly higher estimated mastery probabilities than Brazil, but the estimated probabilities were still relatively low compared with most other education systems. These results suggest that students in these two education systems were less likely to be classified as having mastered the four PISA mathematics content domains in the present diagnostic model.

In North America, Canada showed higher estimated mastery probabilities than the United States across all four domains. In Europe, Finland showed higher estimated mastery probabilities than the United Kingdom and Russia in Change and Relationships, Space and Shape, and Quantity, whereas the United Kingdom showed a higher estimated mastery probability in Uncertainty and Data. Australia showed a relatively high estimated mastery probability in Change and Relationships (0.709), but a lower estimated mastery probability in Uncertainty and Data (0.493). Overall, these results indicate that students' mathematics performance should not be interpreted only as a single overall score. Instead, the model-based mastery probabilities reveal domain-specific diagnostic patterns across education systems, highlighting the content domains in which students are more or less likely to demonstrate mastery.

5.3. Analysis of knowledge states

Based on the posterior classification results, the five most frequent knowledge states within each education system were identified, and the corresponding proportions were calculated. As shown in Table 5, the most frequent knowledge state for B-S-J-G China, Singapore, Japan, Canada, the United Kingdom, Finland, and Australia was 1111. This indicates that, in these education systems, the largest proportion of students were classified as having mastered all four diagnostic attributes. The corresponding proportions were 0.381, 0.568, 0.464, 0.442, 0.395, 0.363, and 0.442, respectively. In contrast, the most frequent knowledge state in the remaining education systems was 0000, indicating that the largest proportion of students in these systems were classified as not

having mastered any of the four diagnostic attributes. These findings suggest substantial variation in the distribution of knowledge states across education systems. The second most frequent knowledge state in Brazil was 1101, whereas that in Colombia was 1110. In Brazil, the knowledge state 1101 indicates that 10.8% of students were classified as having mastered Change and Relationships, Space and Shape, and Uncertainty and Data, but not Quantity. In Colombia, the knowledge state 1110 indicates that 12.6% of students were classified as having mastered Change and Relationships, Space and Shape, and Quantity, but not Uncertainty and Data. These patterns suggest that the domains of Quantity and Uncertainty and Data may represent relatively weaker areas in the corresponding model-based diagnostic profiles for these two education systems. The distribution of the remaining frequent knowledge states further suggests domain-specific differences in mastery patterns. Across the eleven education systems, Change and Relationships appeared more frequently as a mastered domain in the common knowledge-state profiles, suggesting that students were more likely to be classified as having mastered this domain than some of the other domains. By contrast, Uncertainty and Data was often absent from otherwise higher-mastery knowledge states, suggesting that this domain may be a relatively challenging area in the present diagnostic classification. In addition, Japan and Russia showed relatively lower proportions of students classified as having mastered Space and Shape in some of their frequent knowledge-state profiles, with proportions of 0.053 and 0.072, respectively. These results should be interpreted as cross-sectional knowledge-state distributions rather than as evidence of students' longitudinal movement among knowledge states.

5.4. Cross-Sectional Mastery-Order Patterns

Figure 2 shows the cross-sectional mastery-order patterns derived from knowledge-state distributions across education systems. The red path denotes the most frequent mastery-order path,

whereas the black paths denote secondary mastery-order paths with relatively high proportions. Multiple black paths indicate that several secondary mastery-order paths have similar proportions. These paths should be understood as descriptive arrangements of knowledge states from lower to higher levels of content-domain mastery, rather than as actual developmental transitions. Taking B-S-J-G China as an example, the lowest mastery level is represented by the knowledge state 0000. At the second level, the knowledge state 1000 appeared more frequently than 0100, 0010, and 0001, indicating that Change and Relationships was more frequently present as the first mastered domain in the model-based knowledge-state distribution. At the third level, the two knowledge states with relatively high proportions were 1100 and 0011. The knowledge state 1100 includes mastery of Change and Relationships and Space and Shape, and therefore has an inclusion relationship with 1000. In this descriptive sense, a mastery-order path from 1000 to 1100 can be identified. Similarly, at the fourth level, the knowledge state 1110 includes mastery of Change and Relationships, Space and Shape, and Quantity. Thus, the dominant cross-sectional mastery-order path for B-S-J-G China can be summarized as: 0000→1000→1100→1110→1111. This pattern indicates that, within the cross-sectional diagnostic classification, Change and Relationships and Space and Shape appeared earlier than Quantity and Uncertainty and Data in the dominant knowledge-state arrangement. However, this should not be interpreted as evidence that individual students actually mastered these domains in this temporal order. As shown in Figure 2, most education systems displayed the same dominant cross-sectional mastery-order pattern: 0000→1000→1100→1110→1111. This pattern was observed as the dominant red path in all education systems except Japan. It suggests that, in the model-based knowledge-state distributions, Change and Relationships was more frequently present in lower-level mastery profiles, followed by Space and Shape, Quantity, and finally Uncertainty and Data. However, this ordering reflects

the relative frequency of knowledge-state combinations in the cross-sectional sample and should not be interpreted as a longitudinal sequence of individual student development. Japan showed a different dominant cross-sectional mastery-order pattern: 0000→1000→1010→1011→1111. This pattern suggests that Space and Shape was less frequently present in the intermediate mastery states for Japan than in the dominant patterns observed in most other education systems. In other words, the Japan pattern should not be interpreted as showing that Japanese students "took longer" to master Space and Shape; rather, it indicates that Space and Shape was less commonly included in the intermediate knowledge-state combinations in the cross-sectional diagnostic classification. Australia and the United Kingdom showed similar cross-sectional mastery-order structures, and Russia and Canada also showed similar mastery-order structures. In these education systems, three relatively frequent paths from 0000 to 1111 were identified. These multiple paths suggest that students' model-based mastery profiles were not concentrated in a single dominant arrangement, but instead showed several common combinations of mastered and non-mastered content domains. Such patterns indicate heterogeneity in the distribution of diagnostic profiles within these education systems. Brazil displayed four relatively frequent cross-sectional mastery-order paths from 0000 to 1111. These paths suggest that the Brazilian sample showed multiple common arrangements of content-domain mastery rather than a single dominant pattern. The presence of four paths indicates greater heterogeneity in the distribution of model-based knowledge states, with the full-mastery state connected to several relatively frequent intermediate knowledge-state combinations. Colombia showed the simplest branching structure, with two high-frequency cross-sectional mastery-order paths: 0000→1000→1100→1110→1111, and 0000→0100→1100→1110→1111. The first path indicates that Change and Relationships appeared as the first mastered domain before Space and Shape, Quantity, and Uncertainty and Data

were included in the higher-level knowledge states. The second path indicates that Space and Shape appeared as the first mastered domain before the path converged with 1100 and then proceeded to 1110 and 1111. These two paths suggest that, in the Colombia sample, the cross-sectional knowledge-state distribution was more concentrated in a smaller number of mastery-order arrangements than in education systems with more branching patterns. Together, these two paths indicate a relatively concentrated branching structure in the Colombia sample. Singapore showed a more complex branching structure with multiple secondary mastery-order paths. This pattern suggests greater heterogeneity in Singapore students' model-based cognitive diagnostic profiles across the four PISA mathematics content domains. In other words, although Singapore showed high average mastery probabilities, the knowledge-state distribution also indicates multiple combinations of mastered and non-mastered domains among students. Such branching patterns may provide useful diagnostic information about within-system diversity in students' mastery profiles.

These cross-sectional mastery-order patterns summarize how mastery of the four PISA mathematics content domains tends to be combined within each education system. They may help identify domains that appear more frequently in lower-mastery profiles and domains that are more often absent from otherwise higher-mastery profiles. For example, when Uncertainty and Data is frequently absent from higher-mastery knowledge states, this may suggest that data-related and probabilistic reasoning require additional instructional attention. Nevertheless, such interpretations should be made cautiously. The present analysis does not directly model curriculum structure, educational resources, cultural values, or teacher training. Therefore, the observed mastery-order patterns should be treated as descriptive diagnostic evidence and as a basis for generating

hypotheses, rather than as causal evidence about why these patterns differ across education systems.

5.5. Response-Time-Informed Item Diagnostics

Based on the expected a posteriori estimates of the reparameterized DINA item parameters, the parameters were transformed into the traditional slipping and guessing parameters to interpret item-level diagnostic characteristics. To simplify the presentation, this section focuses on three representative education systems: B-S-J-G China in Asia, the United States in North America, and Finland in Europe. The estimated slipping and guessing parameters are shown in Figure 3. For B-S-J-G China, the three items with the lowest estimated slipping probabilities were CM033Q01, CM155Q01, and CM411Q02, with values of 0.105, 0.109, and 0.152, respectively. These results suggest that students classified as having mastered the required diagnostic attributes were less likely to answer these items incorrectly than the other selected items. The three items with the highest estimated guessing probabilities were CM033Q01, CM474Q01, and CM155Q01, with values of 0.615, 0.575, and 0.383, respectively. Items CM033Q01 and CM155Q01 showed both relatively low slipping probabilities and relatively high guessing probabilities, suggesting that these items may be less effective in distinguishing mastery from non-mastery because students classified as non-masters still had relatively high probabilities of answering them correctly.

For the United States, the items CM033Q01, CM474Q01, and CM155Q01 showed relatively low slipping probabilities and relatively high guessing probabilities. This pattern suggests that these items may provide weaker diagnostic separation between students classified as masters and non-masters in the United States sample. For Finland, the three items with the highest estimated slipping probabilities were CM442Q02, CM803Q01, and CM034Q01, with values of 0.460, 0.304, and 0.258, respectively. The same three items also showed relatively low estimated

guessing probabilities, with values of 0.069, 0.098, and 0.199, respectively. This pattern suggests that these items may be more demanding for students classified as masters but less likely to be answered correctly by students classified as non-masters. Across the eleven education systems, items CM033Q01 and CM474Q01 generally showed relatively high guessing probabilities, whereas items CM803Q01, CM442Q02, and CM034Q01 generally showed relatively low guessing probabilities. Items CM033Q01, CM155Q01, and CM474Q01 generally showed relatively low slipping probabilities in most education systems. These item-level patterns provide diagnostic information about how different items function in the cognitive diagnosis model. They should not be interpreted simply as item difficulty or as direct evidence of students' cognitive processes; rather, they describe model-based slipping and guessing tendencies conditional on the diagnostic classification. The item discrimination index was also examined to evaluate how strongly each item differentiated students with different levels of latent mastery or ability. Figure 4 displays the item discrimination indices across education systems. In the figure, a larger sector area indicates stronger item discrimination, whereas a smaller sector area indicates weaker item discrimination. Taking B-S-J-G China as an example, items CM474Q01 and CM033Q01 had relatively low discrimination indices, with values of 0.225 and 0.279, respectively. In contrast, items CM442Q02 and CM034Q01 showed relatively stronger discrimination, suggesting that these items provided more information for distinguishing students with different levels of latent mastery or ability in the B-S-J-G China sample.

Overall, the response-time-informed item diagnostics complement the attribute mastery and knowledge-state analyses by providing item-level evidence about slipping, guessing, discrimination, and response behavior. These results help explain why items linked to the same mathematics content domain may contribute differently to diagnostic classification. They also

provide useful information for assessment development, because items with high slipping, high guessing, or low discrimination may require further review in terms of cognitive demand, wording, or diagnostic alignment.

6. Discussion

Using data from the PISA 2015 computer-based mathematics assessment, this study examined students' cognitive diagnostic profiles across eleven education systems. Rather than focusing only on overall achievement scores or country-level rankings, the study provided a more fine-grained diagnostic interpretation of students' mastery of four PISA mathematics content domains: Change and Relationships, Space and Shape, Quantity, and Uncertainty and Data. The findings show that cognitive diagnostic modeling can extend the interpretation of PISA data beyond scale scores by estimating students' mastery probabilities and knowledge-state distributions across clearly defined mathematics content domains.

A key contribution of this study is that it links the PISA mathematics framework with cognitive diagnostic assessment. In the official PISA framework, Change and Relationships, Space and Shape, Quantity, and Uncertainty and Data are defined as mathematics content domains. In this study, these domains were operationalized as diagnostic attributes in the cognitive diagnosis model. This made it possible to estimate students' mastery probabilities for each domain and to construct cognitive diagnostic profiles at both the individual and education-system levels. Compared with traditional score-based summaries, these profiles provide more detailed information about the specific content domains in which students may need targeted support. The results revealed substantial cross-national differences in students' mastery probabilities across the four mathematics content domains. Some education systems showed relatively high mastery probabilities across all domains, whereas others displayed more pronounced weaknesses in

specific domains. These findings suggest that students' mathematics performance should not be understood as a single undifferentiated ability. Instead, students may exhibit different mastery patterns across different content domains. For example, relatively high mastery of Change and Relationships in several systems suggests stronger familiarity with algebraic, functional, or relational reasoning, whereas lower mastery of Uncertainty and Data indicates that probabilistic thinking, data interpretation, and reasoning under uncertainty may remain challenging for many students. The knowledge-state results further enrich this interpretation by showing how different combinations of mastered and non-mastered content domains are distributed within each education system. In several higher-performing systems, the most common knowledge state was full mastery of the four diagnostic attributes, whereas in lower-performing systems, a larger proportion of students were classified into lower-mastery knowledge states. These results show that two education systems with similar overall performance may still differ in the internal composition of students' mastery profiles, which has important implications for targeted instructional support. The cross-sectional mastery-order patterns were inferred from the distribution of knowledge states at a single assessment point. For example, a pattern such as 0000 → 1000 → 1100 → 1110 → 1111 indicates that certain combinations of mastered content domains occurred more frequently and could be arranged from lower to higher levels of content-domain mastery. These patterns provide descriptive diagnostic information about the organization of students' mastery profiles, while future longitudinal or repeated-measure studies would be needed to examine actual learning progressions over time. The incorporation of response-time information provides another contribution. In computer-based assessments, response times offer supplementary evidence about students' response behavior and item characteristics. In this study, response time was not treated as an independent cultural variable, but as diagnostic information for interpreting item-level

response processes. By jointly considering response accuracy and response time, the analysis provided a more nuanced account of slipping, guessing, and discrimination. This is useful because students with the same response score may differ substantially in the time spent on an item and, therefore, in the likely response process underlying the observed answer. The item-level results also have implications for assessment development. Items with high slipping probabilities may indicate that students who have mastered the relevant content domain still make errors, possibly because the item is cognitively demanding, linguistically complex, or vulnerable to misunderstanding. Items with high guessing probabilities may be less effective in distinguishing mastery from non-mastery. Similarly, discrimination parameters show how effectively items differentiate students with different levels of latent mastery or ability. These diagnostics can help test developers identify which items contribute most strongly to diagnostic classification and which items may require revision. The cross-national differences observed in this study should be interpreted cautiously and in relation to broader educational contexts. Differences in mastery probabilities and knowledge-state distributions may be associated with curriculum organization, instructional emphasis, teacher preparation, assessment traditions, and educational resources. However, the present study does not directly model curriculum documents, classroom instruction, school resources, or cultural values. Therefore, the observed differences should not be interpreted as direct causal evidence that one factor explains why a particular education system shows stronger mastery in a given domain. Similarly, response-time-informed item differences should not be treated as direct evidence of cultural differences in test-taking behavior without further supporting analyses.

Nevertheless, the practical implications of the study are meaningful. For curriculum designers, the diagnostic results can help identify mathematics content domains that may require

greater emphasis or more appropriate instructional sequencing. For teachers, mastery probabilities and knowledge-state profiles may provide more actionable information than total scores because they indicate the specific domains students are likely to have mastered or not mastered. For assessment developers, slipping, guessing, discrimination, and response-time information can support item evaluation and test refinement. For policy makers, cross-national diagnostic profiles can inform more targeted and evidence-based discussions about mathematics education, especially when interpreted together with contextual information about curriculum, instruction, and educational resources.

7. Limitations

Although this study conducted an in-depth analysis of PISA data from eleven education systems with the help of cognitive diagnosis theory and RTs as auxiliary information, there are still many areas to be improved. First, the present study focused on four PISA mathematics content domains-Change and Relationships, Space and Shape, Quantity, and Uncertainty and Data-which were operationalized as diagnostic attributes in the cognitive diagnosis model. Other important aspects of mathematical literacy, such as mathematical processes, reasoning, problem solving, and contextual dimensions, were not included in the current diagnostic model. This decision was made because the publicly available item-level information for the selected PISA items did not provide sufficient evidence to construct a defensible Q-matrix for these additional dimensions. Future research could extend the diagnostic framework by incorporating additional dimensions when reliable item–attribute mapping information is available. Second, the current analysis was based on a limited set of nine common computer-based mathematics items. This restriction was necessary to ensure that all students in the analytic sample responded to the same items and that comparisons across education systems were based on a common assessment basis. However, the

small number of items limits the breadth of content-domain coverage and may affect the precision of attribute mastery classification. Future studies could examine larger item sets or use designs that allow broader content coverage while preserving comparability across education systems. Third, each selected item in the present analysis was linked to one primary diagnostic attribute. This single-attribute specification was adopted to maintain conceptual clarity and estimation stability. However, mathematical problem solving often involves multiple content domains and processes. Future research could examine the feasibility of using multi-attribute Q-matrices or more general cognitive diagnosis models, such as G-DINA or LCDM, when the item design and available documentation support such specifications. In addition, the accuracy of the Q-matrix is a critical assumption in cognitive diagnosis modeling. If expert-defined Q-matrices are misspecified, posterior classification and item parameter estimates may be affected. Future work may therefore consider data-driven Q-matrix validation or refinement methods that use both item responses and response-time information. Fourth, the present study used cross-sectional PISA 2015 data. Therefore, the knowledge-state patterns reported in this study cannot be interpreted as longitudinal learning trajectories or as direct evidence of students' cognitive development over time. The cross-sectional mastery-order patterns summarize how different combinations of mastered and non-mastered content domains are distributed at a single assessment point. Although these patterns can generate hypotheses about possible instructional sequences or content-domain dependencies, longitudinal assessment data or repeated-measure designs would be needed to examine actual developmental progressions. Fifth, the present model did not include student-, school-, or system-level background variables, such as socioeconomic status, curriculum exposure, instructional practices, educational resources, or teacher professional development. This choice was made to keep the analysis focused on the cognitive diagnosis and response-time modeling

framework. However, educational outcomes are shaped by broader social, institutional, and instructional contexts. Curriculum organization, learning opportunities, resource allocation, school-level conditions, and teacher support may all be related to students' mastery patterns across mathematics content domains (Akiba, LeTendre, & Scribner, 2007; Heitink et al., 2016; Heyneman & Loxley, 1983; Schmidt & Prawat, 2006). Because these factors were not directly modeled, the present study cannot explain why specific education systems showed particular diagnostic patterns. Future research could integrate cognitive diagnosis models with multilevel modeling or structural equation modeling to examine how individual-, school-, and system-level factors are associated with attribute mastery and knowledge-state distributions. Finally, response time information was used as supplementary diagnostic information rather than as a direct measure of students' cognitive processes or cultural differences in test-taking behavior. Although response times can provide useful information about response behavior and item functioning, they should be interpreted cautiously. Similar response times may reflect different response strategies, motivation levels, or item-processing behaviors. Future studies could combine response-time data with additional process data, such as action logs, revisits, or item-level interaction records, to better understand students' response processes in computer-based assessments.

8. Conclusions

This study applied a joint cognitive diagnosis and response-time modeling framework to the PISA 2015 computer-based mathematics assessment data from eleven education systems. The study focused on four PISA mathematics content domains—Change and Relationships, Space and Shape, Quantity, and Uncertainty and Data—which were operationalized as diagnostic attributes in the cognitive diagnosis model. By estimating attribute mastery probabilities, knowledge-state distributions, cross-sectional mastery-order patterns, and response-time-informed item diagnostics,

the study provided a more fine-grained interpretation of students' mathematics performance than can be obtained from overall scores alone.

The findings make three main contributions. First, the study shows that cognitive diagnostic modeling can decompose students' mathematics performance into domain-specific mastery profiles. This helps identify the content domains in which students are more or less likely to demonstrate mastery within each education system. Second, the analysis of knowledge-state distributions provides information about how different combinations of mastered and non-mastered content domains are distributed across education systems. The resulting cross-sectional mastery-order patterns offer descriptive diagnostic information about the relative ordering of content-domain mastery. Third, the response-time-informed item diagnostics provide additional item-level evidence about slipping, guessing, discrimination, item time intensity, and response behavior. These results help explain why items linked to similar content domains may contribute differently to diagnostic classification.

The study also highlights the practical value of integrating cognitive diagnosis and response-time modeling in international large-scale assessments. For curriculum designers, domain-specific mastery probabilities and knowledge-state distributions can help identify content domains that may require greater instructional emphasis or more appropriate sequencing. For teachers, cognitive diagnostic profiles may provide more actionable information than total scores by indicating which content domains students are likely to have mastered or not mastered. For assessment developers, slipping, guessing, discrimination, and response-time information can help evaluate item functioning and improve the diagnostic quality of computer-based assessments. For policy makers, cross-national diagnostic profiles can support more targeted and evidence-informed

discussions about mathematics education, especially when interpreted together with contextual information about curriculum, instruction, and educational resources.

At the same time, the conclusions of this study should be interpreted cautiously. The analysis was based on cross-sectional PISA data and a limited set of common items. Therefore, it cannot establish causal relationships among curriculum, instruction, cultural context, educational resources, and students' mastery patterns. Nor can it provide direct evidence of students' longitudinal cognitive development. Future research using longitudinal data, broader item sets, process data, and multilevel contextual information will be needed to further examine how students' mastery of mathematics content domains develops over time and how educational contexts are associated with diagnostic profiles.

In conclusion, this study demonstrates that cognitive diagnosis and response-time modeling can enrich the interpretation of PISA computer-based mathematics assessment data. By moving beyond overall performance comparisons, the proposed framework provides a more detailed picture of students' content-domain mastery, knowledge-state distributions, cross-sectional mastery-order patterns, and item-level diagnostic characteristics. This framework offers a useful methodological and practical basis for improving the interpretability of international large-scale assessments and for supporting more diagnostic and evidence-informed approaches to mathematics education.

References

Akiba, M., LeTendre, G. K., & Scribner, J. P. (2007). Teacher quality, opportunity gap, and national achievement in 46 countries. *Educational Researcher*, *36(7)*, 369–387.

Bolsinova, M., De Boeck, P., & Tijmstra, J. (2017). Modelling conditional dependence between response time and accuracy. *Psychometrika*, *82*, 1126-1148.

Bolsinova, M., & Tijmstra, J. (2018). Improving precision of ability estimation: Getting more from response times. *British Journal of Mathematical and Statistical Psychology*, *71*, 13-38

Brooks, S. & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*, 434-55.

De Boeck, P., & Jeon, M. (2019). An overview of models for response times and processes in cognitive tests? *Frontiers in Psychology*, *10*, 102.

DeCarlo, L. T. (2011). On the analysis of fraction subtraction data: The DINA model, classification, latent class sizes, and the Q-matrix. *Applied Psychological Measurement*, *35*, 8-26.

de la Torre, J. (2008). An empirically-based method of Q-matrix validation for the DINA model: Development and applications. *Journal of Educational Measurement*, *45*, 343-362

de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, *34*, 115-130.

de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, *76*, 179-199.

de la Torre, J., & Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, *69*, 333-353.

Fischman, G. E., Topper, A. M., Silova, I., Goebel, J., & Holloway, J. L. (2019). Examining the influence of international large-scale assessments on national education policies. *Journal of Education Policy*, *34*(4), 470-499.

Fox, J.-P., Klein Entink, R. H., & van der Linden, W. J. (2007). Modeling of responses and response times with the package CIRT. *Journal of Statistical Software*, *20*(7), 1–14.

Gravemeijer, K., Bowers, J., & Stephan. M. A. (2003). Hypothetical learning trajectory in measurement and flexible arithmetic. *Journal for Research in Mathematics Education. Monograph*, *12*, 51-66.

Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, *26*, 301-321.

Hanushek, E. A., & Woessmann, L. (2011). How much do educational outcomes matter in OECD countries? *Economic Policy*, *26*(67), 427-491.

Hartz, S. M. (2002). *A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality*. (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign, Champaign, IL.

Heitink, M. C., van der Kleij, F. M., Veldkamp, B. P., Schildkamp, K., & Kippers, W. B. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational Research Review*, *17*, 50–62.

Henson, R., Templin, J., & Willse, J. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, *74*, 191-210.

Heyneman, S. P., & Loxley, W. A. (1983). The effect of primary-school quality on academic achievement across twenty-nine high-and low-income countries. *American Journal of Sociology*, *88(6)*, 1162–1194.

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25, 258-272.

Klein Entink, R. H., Fox, J.-P., & van der Linden, W. J. (2009). A multivariate multilevel approach to the modeling of accuracy and speed of test takers. *Psychometrika*, *74*(1), 21–48.

Lestari, Th. V. D., & Marsigit. (2020). Hypothetical learning trajectory and students' understanding of the concepts of the arithmetic sequence. *Journal of Physics Conference Series*, *1581*, 1-8.

Li, J., Xue, E., & Guo, S. (2025). The effects of PISA on global basic education reform: a systematic literature review. *Humanities and Social Sciences Communications*, *12*(1), 1-15.

Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.

Lu, J., Wang, C, Zhang, J., W., & Tao, J. (2020). A mixture model for responses and response times with a higher-order ability structure to detect rapid guessing behaviour. *British Journal of Mathematical and Statistical Psychology*, *73*, 261-288.

Maris, E. (1993). Additive and multiplicative models for gamma distributed variables, and their application as psychometric models for response times. *Psychometrika*, *58*, 445-469.

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, *64*, 187-212.

Merritt, J. D., Krajcik, J., & Shwartz, Y. (2008). *Development of a learning progression for the particle model of matter*. In Proceedings of the 8th International Conference for the Learning Sciences-Volume 2 (pp. 75-81). Utrecht, The Netherlands: International Society of the Learning Sciences.

Nortvedt, G. A. (2018). Policy impact of PISA on mathematics education: The case of Norway. *European Journal of Psychology of Education*, *33*(3), 427-444.

Nyimbili , F., & Nyimbili , L. (2024). Types of purposive sampling techniques with their examples and application in qualitative research studies. *British Journal of Multidisciplinary and Advanced Studies, 5(1)*, 90–99.

OECD. (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris: OECD Publishing.

OECD. (2016). *PISA 2015 assessment and analytical framework: Science, reading, mathematics and financial literacy*. Paris: OECD Publishing.

Patton, M. Q. (2002). *Qualitative research and evaluation methods (Vol. 3)*. Sage.

Plummer, M. (2015). *JAGS Version 4.0.0 user manual*. Lyon, France. Retrieved from http://sourceforge.net/projects/mcmc-jags/.

Qian, H., Staniewska, D., Reckase, M., & Woo, A. (2016). Using response time to detect item preknowledge in computer-based licensure examinations. *Educational Measurement: Issues and Practice*, *35*(1), 38-47.

Qiao, X., & Jiao, H. (2021). Explanatory cognitive diagnostic modeling incorporating response times. *Journal of Educational Measurement*, *58*, 564-585.

Ranger, J., & Ortner, T. (2012). A latent trait model for response times on tests employing the proportional hazards model. *British Journal of Mathematical and Statistical Psychology*, *65*, 334-349.

Rouder, J. N., Sun, D., Speckman, P. L., Lu, J., & Zhou, D. (2003). A hierarchical Bayesian statistical framework for response time distributions. *Psychometrika*, *68*, 589-606.

Santos, Í., & Centeno, V. G. (2023). Inspirations from abroad: The impact of PISA on countries' choice of reference societies in education. Compare: *A Journal of Comparative and International Education*, *53*(2), 269-286.

Scheiblechner, H. (1979). Specific objective stochastic latency mechanisms. *Journal of Mathematical Psychology*, *19*, 18-38.

Schmidt, W. H., & Prawat, R. S. (2006). Curriculum coherence and national control of education: Issue or non‐issue? *Journal of Curriculum Studies*, *38(6)*, 641–658.

Schnipke, D. L., & Scrams, D. J. (1997). Modeling response times with a two-state mixture model: A new method of measuring speededness. *Journal of Educational Measurement*, *34*, 213-232.

Simola, H. (2005). The Finnish miracle of PISA: historical and sociological remarks on teaching and teacher education. *Comparative Education, 41(4)*, 455–470.

Songer, N. B., Kelcey, B., & Gotwals, A. W. (2009). How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning about biodiversity. *Journal of Research in Science Teaching*, *7*, 611-631.

Su, Y.-S., & Yajima, M. (2015). *R2jags: Using R to run "JAGS"* (R package version 0.5-7). Retrieved from http://CRAN.R-project.org / package = R2jags

Tatsuoka, C. (2002). Data analytic methods for latent partially ordered classification models. *Journal of the Royal Statistical Society, Series C: Applied Statistics*, *51*, 337-350.

Tatsuoka, K. K. (1984). *Analysis of errors in fraction addition and subtraction problems*. Computer-Based Education Research Laboratory, University of Illinois at Urbana-Champaign, Champaign.

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, *11*, 287-305.

van der Linden, W. J. (2006). A lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics*, *31*, 181-204.

van der Linden, W. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika*, *72*, 297-308.

von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, *61*, 287-307.

von Davier, M. (2014). The DINA model as a constrained general diagnostic model: Two variants of a model equivalency. *British Journal of Mathematical and Statistical Psychology*, *67*, 49-71.

Wang, C., Chang, H. H., & Douglas, J. A. (2013). The linear transformation model with frailties for the analysis of item response times. *British Journal of Mathematical and Statistical Psychology*, *66*, 144-168.

Wang, C., Fan, Z., Chang, H.-H., & Douglas, J. (2013). A semiparametric model for jointly analyzing response times and accuracy in computerized testing. *Journal of Educational and Behavioral Statistics*, *38*, 381-417.

Wang, C., & Xu, G. (2015). A mixture hierarchical model for response times and response accuracy. *British Journal of Mathematical and Statistical Psychology*, *68*, 456-477.

Wise, S. L., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education*, *18*, 163-183.

Zhan, P., Jiao, H., & Liao, D. (2018a). Cognitive diagnosis modelling incorporating item response times. *British Journal of Mathematical and Statistical Psychology*, *71*, 262-286.

Zhan, P., Liao, M., & Bian, Y. (2018b). Joint testlet cognitive diagnosis modeling for paired local item dependence in response times and response accuracy. *Frontiers in Psychology*, *9*, 607.

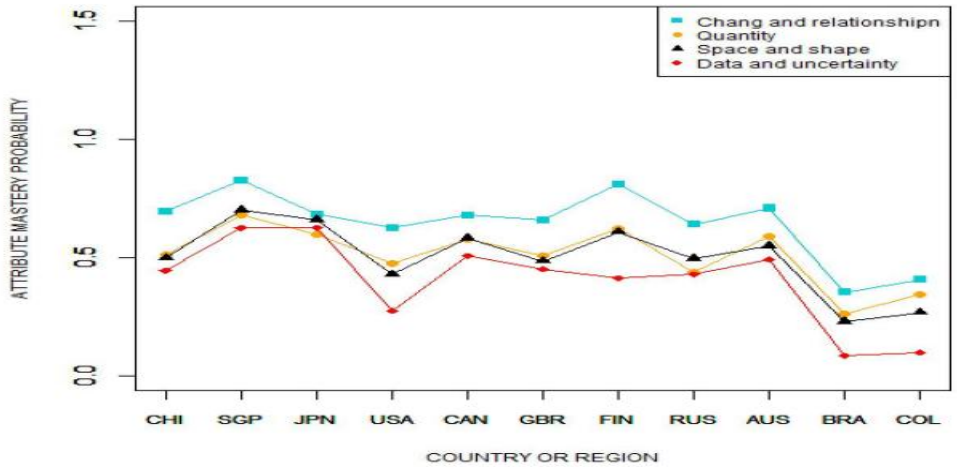
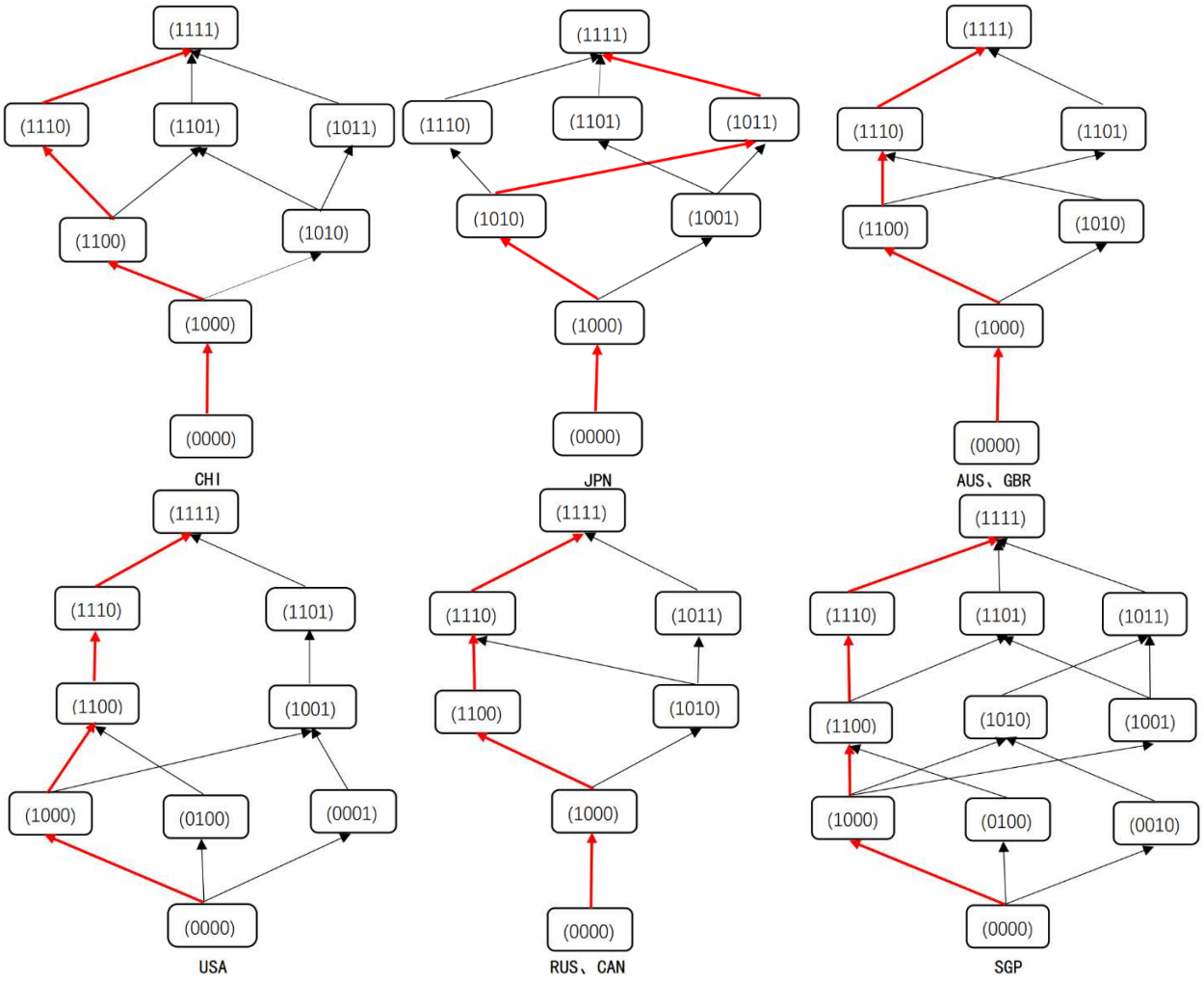


Figure 1: Average posterior attribute mastery probabilities across 11 education systems.



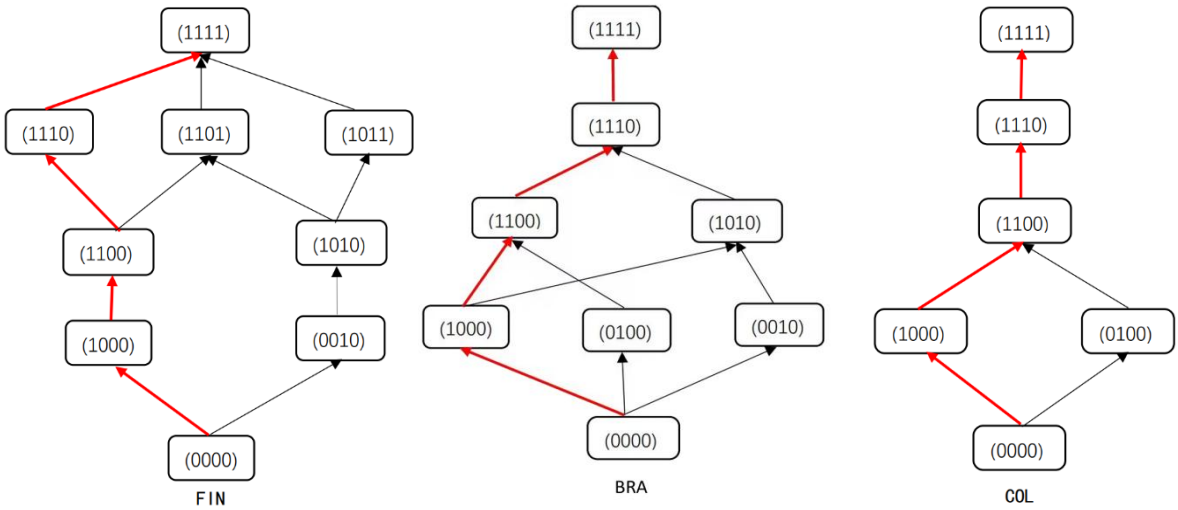


Figure 2: Cross-sectional mastery-order patterns derived from knowledge-state distributions across education systems.

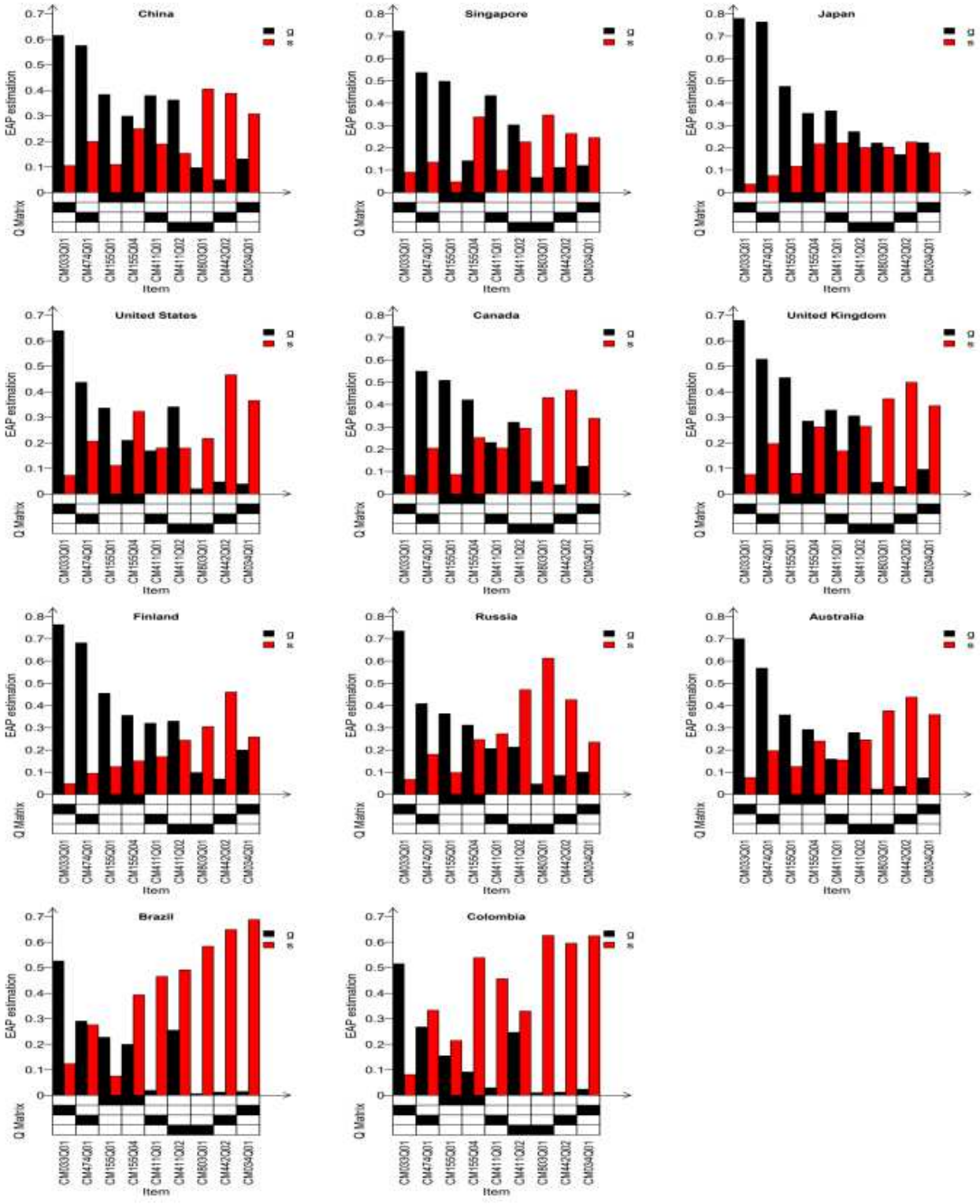


Figure 3: EAP estimates of the slipping and guessing parameters for different education systems.

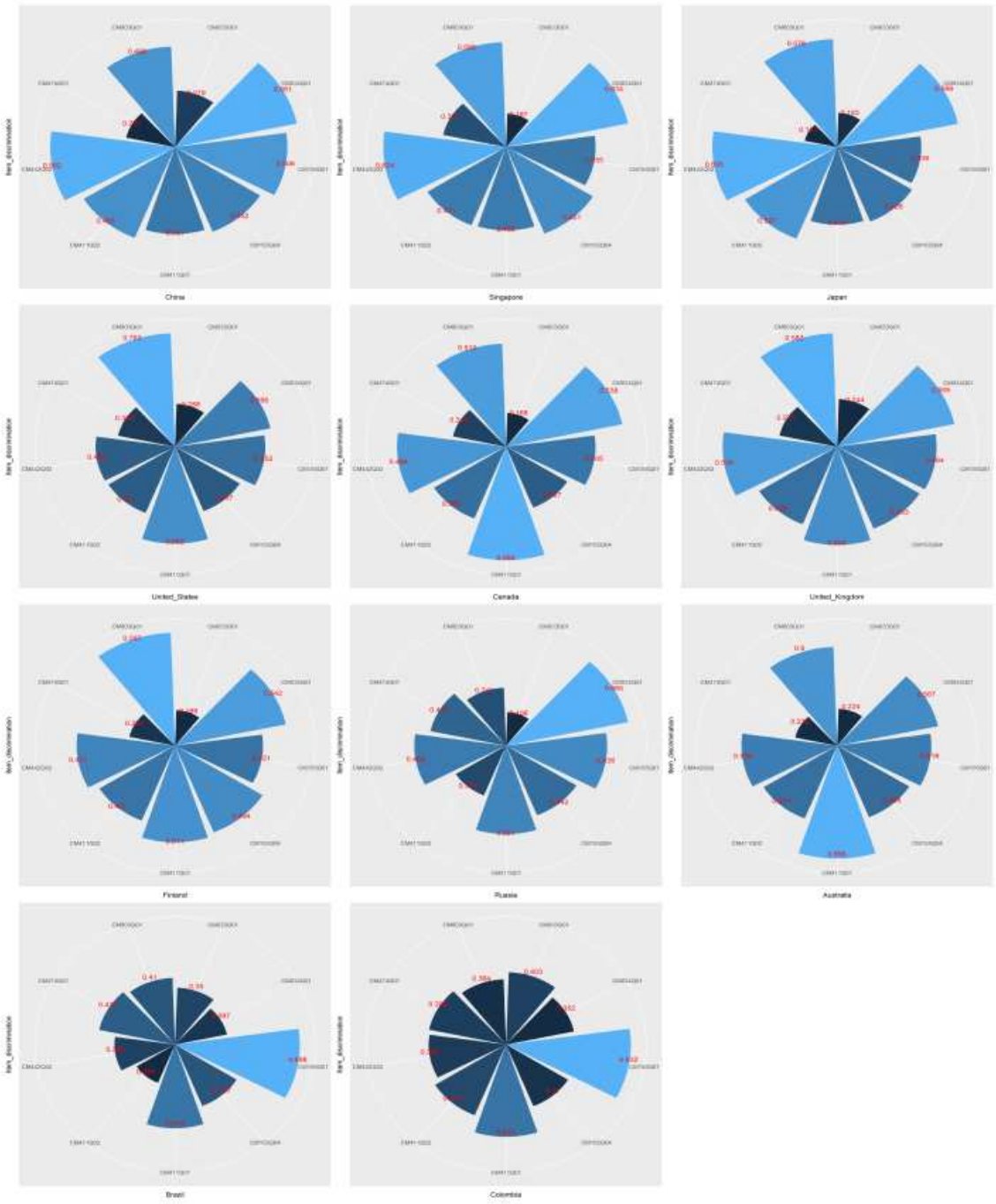


Figure 4: Item discriminations for nine test items in different education systems.

Table 1: PISA mathematics content domains operationalized as diagnostic attributes.

Attribute	Definition
Change and Relationships (α_1)	Involves the use of mathematical language, such as algebraic expressions, equations, functions, and inequalities, to describe relationships among quantities and graphical representations.
Space and Shape (α_2)	Involves spatial and geometric relationships among points, lines, planes, and shapes, as well as the mental rotation and visual representation of figures.
Quantity (α_3)	Involves the quantification of attributes of objects, understanding different representations of quantities, and judging, interpreting, and reasoning about quantitative information.
Uncertainty and Data (α_4)	Involves probability, chance, data representation, data interpretation, and the evaluation of information involving uncertainty.

Table 2: Q-matrix for the nine selected PISA mathematics items.

Item	Attribute			
	Change and Relationships (α_1)	Space and Shape (α_2)	Quantity (α_3)	Uncertainty and Data (α_4)
CM033Q01	0	1	0	0
CM474Q01	0	0	1	0
CM155Q01	1	0	0	0
CM155Q04	1	0	0	0
CM411Q01	0	0	1	0
CM411Q02	0	0	0	1
CM803Q01	0	0	0	1
CM442Q02	0	0	1	0
CM034Q01	0	1	0	0

Table 3: Number of nine items correctly answered by students from different education systems.

China (CHI, 1067)				Singapore (SGP, 687)				Japan (JPN, 731)			
Total	Frequency	Percent	Cumulative percent	Total	Frequency	Percent	Cumulative percent	Total	Frequency	Percent	Cumulative percent
0	16	1.50%	1.50%	0	10	1.50%	1.50%	0	4	0.60%	0.60%
1	48	4.50%	6.00%	1	17	2.50%	4.00%	1	12	1.60%	2.20%
2	91	8.50%	14.50%	2	25	3.60%	7.60%	2	22	3.00%	5.20%
3	129	12.10%	26.60%	3	63	9.20%	16.80%	3	58	7.90%	13.10%
4	148	13.90%	40.50%	4	70	10.20%	27.00%	4	74	10.10%	23.20%
5	178	16.70%	57.20%	5	105	15.30%	42.30%	5	111	15.20%	38.40%
6	173	16.20%	73.40%	6	119	17.30%	59.60%	6	117	16.00%	54.40%
7	135	12.60%	86.00%	7	115	16.70%	76.30%	7	132	18.10%	72.50%
8	94	8.80%	94.80%	8	101	14.70%	91.00%	8	134	18.30%	90.80%
9	55	5.20%	100.00%	9	62	9.00%	100.00%	9	67	9.20%	100.00%
United States (USA, 567)				Canada (CAN, 2021)				United Kingdom (GBR, 1396)			
Total	Frequency	Percent	Cumulative percent	Total	Frequency	Percent	Cumulative percent	Total	Frequency	Percent	Cumulative percent
0	16	2.80%	2.80%	0	19	0.90%	0.90%	0	22	1.60%	1.60%
1	53	9.30%	12.10%	1	64	3.20%	4.10%	1	69	4.90%	6.50%
2	65	11.50%	23.60%	2	169	8.40%	12.50%	2	137	9.80%	16.30%
3	78	13.80%	37.40%	3	242	12.00%	24.50%	3	187	13.40%	29.70%
4	95	16.80%	54.20%	4	320	15.80%	40.30%	4	212	15.20%	44.90%
5	87	15.30%	69.50%	5	342	16.90%	57.20%	5	211	15.10%	60.00%
6	58	10.20%	79.70%	6	329	16.30%	73.50%	6	197	14.10%	74.10%
7	60	10.60%	90.30%	7	268	13.30%	86.80%	7	171	12.30%	86.40%
8	37	6.50%	96.80%	8	184	9.10%	95.90%	8	134	9.60%	96.00%
9	18	3.20%	100.00%	9	84	4.10%	100.00%	9	56	4.00%	100.00%
Finland (FIN, 603)				Russia (RUS, 571)				Australia (AUS, 1422)			
Total	Frequency	Percent	Cumulative percent	Total	Frequency	Percent	Cumulative percent	Total	Frequency	Percent	Cumulative percent
0	3	0.50%	0.50%	0	9	1.60%	1.60%	0	24	1.70%	1.70%
1	17	2.80%	3.30%	1	34	6.00%	7.60%	1	81	5.70%	7.40%
2	27	4.50%	7.80%	2	65	11.40%	19.00%	2	122	8.60%	16.00%
3	52	8.60%	16.40%	3	96	16.70%	35.70%	3	184	12.94%	28.94%
4	86	14.30%	30.70%	4	93	16.20%	51.90%	4	214	15.00%	43.94%
5	104	17.20%	47.90%	5	98	17.10%	69.00%	5	201	14.14%	58.08%
6	99	16.40%	64.30%	6	73	12.70%	81.70%	6	205	14.42%	72.50%
7	110	18.20%	82.50%	7	56	9.70%	91.40%	7	169	11.90%	84.40%
8	69	11.50%	94.00%	8	35	6.60%	98.00%	8	158	11.10%	95.50%
9	36	6.00%	100.00%	9	12	2.00%	100.00%	9	64	4.50%	100.00%
Brazil (BRA, 1365)				Colombia (COL, 1101)				Saudi Arabia (ARB, 1387)			
Total	Frequency	Percent	Cumulative percent	Total	Frequency	Percent	Cumulative percent	Total	Frequency	Percent	Cumulative percent
0	151	11.16%	11.16%	0	120	10.90%	10.90%	0	90	6.50%	6.50%
1	274	20.07%	31.23%	1	235	21.30%	32.20%	1	210	15.10%	21.60%
2	314	23.00%	54.23%	2	253	22.90%	55.10%	2	239	17.20%	38.80%
3	253	18.50%	72.73%	3	182	16.50%	71.60%	3	256	18.50%	57.30%
4	179	13.11%	84.84%	4	137	12.30%	83.90%	4	188	13.60%	70.90%
5	92	6.73%	92.57%	5	81	7.40%	91.30%	5	159	11.50%	82.40%
6	40	2.90%	95.47%	6	48	4.40%	95.70%	6	112	8.00%	90.50%
7	35	2.56%	98.03%	7	27	2.60%	98.30%	7	63	4.50%	95.00%
8	23	1.68%	99.71%	8	13	1.20%	99.50%	8	46	3.30%	98.30%
9	4	0.29%	100.00%	9	5	0.50%	100.00%	9	24	1.70%	100.00%

Table 4: Average posterior mastery probabilities for the four diagnostic attributes across 11 education systems.

	Attribute			
Education system	Change and Relationships	Space and Shape	Quantity	Uncertainty and Data
B-S-J-G China	0.695	0.512	0.499	0.445
Singapore	0.826	0.679	0.701	0.625
Japan	0.682	0.599	0.660	0.625
United States	0.626	0.477	0.428	0.275
Canada	0.679	0.578	0.582	0.507
United Kingdom	0.658	0.509	0.485	0.452
Finland	0.801	0.623	0.608	0.412
Russia	0.640	0.440	0.496	0.429
Australia	0.709	0.589	0.550	0.493
Brazil	0.353	0.264	0.230	0.085
Colombia	0.407	0.342	0.268	0.099

Table 5: Five most frequently observed knowledge states in each education system.

Education system	Most frequent knowledge state		Second most frequent knowledge state		Third most frequent knowledge state		Fourth most frequent knowledge state		Fifth most frequent knowledge state	
	State	Proportion	State	Proportion	State	Proportion	State	Proportion	State	Proportion
B-S-J-G China	(1111)	0.381	(0000)	0.279	(1000)	0.147	(1110)	0.066	(1100)	0.034
Singapore	(1111)	0.568	(0000)	0.160	(1000)	0.076	(1110)	0.052	(1011)	0.038
Japan	(1111)	0.464	(0000)	0.182	(1000)	0.078	(1011)	0.053	(1010)	0.034
United States	(0000)	0.326	(1111)	0.240	(1000)	0.125	(1110)	0.109	(1100)	0.081
Canada	(1111)	0.442	(0000)	0.263	(1000)	0.082	(1110)	0.068	(1010)	0.025
United Kingdom	(1111)	0.395	(0000)	0.300	(1000)	0.123	(1110)	0.046	(1100)	0.045
Finland	(1111)	0.363	(0000)	0.156	(1000)	0.143	(1110)	0.134	(1100)	0.065
Russia	(0000)	0.294	(1111)	0.266	(1000)	0.091	(1011)	0.072	(1110)	0.070
Australia	(1111)	0.442	(0000)	0.251	(1000)	0.115	(1110)	0.075	(1100)	0.030
Brazil	(0000)	0.650	(1101)	0.108	(1111)	0.076	(1000)	0.054	(1100)	0.044
Colombia	(0000)	0.534	(1110)	0.126	(1000)	0.085	(1100)	0.082	(1111)	0.080

Note: The four digits in each knowledge state correspond, in order, to Change and Relationships, Space and Shape, Quantity, and Uncertainty and Data. A value of 1 indicates estimated mastery, and a value of 0 indicates estimated non-mastery.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- realdataPISA2015.zip