# A Method for Assessing Scientific Literacy Based on Multi-Order Cognitive Diagnosis Models

**Authors:** Zhan Peida, Yu Zhaohui, Li Feiming, Wang Lijun, Peida Zhan

**Date:** 2019-02-14T12:00:26+00:00

## Abstract

Scientific literacy refers to the ability to solve scientific problems and apply scientific concepts as a reflective citizen. For the assessment of scientific literacy in cognitive diagnosis, this paper, based on the PISA 2015 scientific literacy assessment framework, proposes for the first time a third-order latent structure underlying scientific literacy, analyzes the PISA 2015 scientific assessment data using a newly proposed multi-order cognitive diagnosis model, and investigates the psychometric properties of the new model through simulation studies. The results indicate that: (1) the new model can adequately analyze scientific literacy comprising a third-order latent structure; (2) scientific knowledge has the greatest influence on scientific literacy, followed by scientific contexts, with scientific competencies having the least influence; (3) the full Bayesian MCMC algorithm can provide relatively accurate parameter estimates for the new model.

## Full Text

### A Method for Assessing Scientific Literacy Based on Multi-Order Cognitive Diagnosis Models

**Acta Psychologica Sinica** 2019, Vol. 51, No. 5, 1–13
DOI: 10.3724/SP.J.1041.2019.001

**ZHAN Peida; YU Zhaohui; LI Feiming; WANG Lijun**
(College of Teacher Education, Zhejiang Normal University, Jinhua, 321004, China)

#### Abstract

Scientific literacy refers to the capacity of individuals to solve scientific problems and apply scientific concepts as reflective citizens. To achieve objective assessment of scientific literacy within cognitive diagnostic assessment, this study proposes, for the first time, a third-order latent structure of scientific literacy based on the PISA 2015 scientific literacy assessment framework. Using a newly proposed multi-order cognitive diagnosis model, we analyzed PISA 2015 science assessment data and investigated the psychometric properties of the new model through simulation studies.

The results indicate that: (1) the new model can adequately analyze scientific literacy comprising a third-order latent structure; (2) scientific knowledge has the greatest influence on scientific literacy, followed by scientific contexts, with scientific competencies having the smallest effect; and (3) the full Bayesian MCMC algorithm can provide accurate parameter estimation for the new model.

**Keywords:** scientific literacy; cognitive diagnosis; PISA; DINA model
**Classification Number:** B841

---

## 1. Introduction

Scientific and technological advances drive productivity, economic prosperity, and social progress, transforming people's modes of production, lifestyles, and ways of thinking. The rapid development of science and technology imposes new demands on the scientific literacy of every citizen (Ministry of Education of the People's Republic of China, 2017). In essence, scientific literacy refers to the ability of reflective citizens to solve scientific problems and apply scientific concepts. The 2017 *Compulsory Education Primary School Science Curriculum Standards* defines scientific literacy as "understanding necessary scientific and technological knowledge and its impact on society and individuals, knowing basic scientific methods, recognizing the nature of science, establishing scientific thinking, advocating scientific spirit, and possessing the ability to apply them to practical problems and participate in public affairs."

To achieve objective assessment of scientific literacy, the Programme for International Student Assessment (PISA) specified four interdependent dimensions in its 2015 framework: Competencies, Knowledge, Contexts, and Attitudes [Figure 1: see original paper]. This framework requires students to demonstrate their scientific competencies by applying scientific knowledge to solve problems within specific scientific contexts, guided by their scientific attitudes (Liu & Li, 2015). The PISA 2015 assessment framework represents an evolution from the 2006 framework (OECD, 2006), primarily through more detailed classification of the knowledge dimension. The gradual refinement of the assessment framework reflects ongoing reconceptualization of scientific literacy based on practical experience. The PISA 2015 framework currently represents the most recent and operationalizable approach to assessing scientific literacy.

In addition to an operationalizable assessment framework, an appropriate assessment method is equally crucial. Such methods should align with the assessment framework and enable objective and accurate evaluation of scientific literacy.

However, existing research predominantly relies on questionnaire surveys of citizens' or students' scientific literacy (e.g., Roos, 2014; Gao, 2011; Qin & Qian, 2008), which only provide a general understanding of the overall status of scientific literacy. These surveys typically employ self-report methods, which are subjective and susceptible to social desirability bias. Only a few studies have focused on actual assessment of scientific literacy (e.g., Hu, Yang, & Lu, 2012). Beyond methodological limitations, most current research employs outdated theories and analytical methods, primarily classical test theory (e.g., Roos, 2014; Ren, Zhang, & He, 2013), with only occasional use of item response theory (IRT) models (e.g., Hu et al., 2012).

Importantly, although PISA constructed a multidimensional structure for scientific literacy, unidimensional IRT models were used for data analysis (OECD, 2017). This means the analytical model does not match the assessment framework. The primary reason is that PISA focuses on national/economic-level patterns rather than individual participants, making a unidimensional latent trait approach convenient for simplifying overall research complexity. However, when individuals become the assessment focus, more sophisticated methods are required (e.g., Zhan, Jiao, & Liao, 2018). In summary, achieving objective and accurate assessment of scientific literacy under the PISA 2015 framework requires new perspectives and more appropriate assessment methods.

Recent developments in cognitive psychology have revealed that completing tasks often requires coordination of multiple abilities, rendering the unidimensionality assumption in early psychometric models unrealistic (Reckase, 2009; Wang & Chen, 2004; Kang & Xin, 2010; Zhan, Wang, & Wang, 2013). Moreover, beyond simple total scores, researchers seek richer information from actual response patterns to make more objective evaluations and provide remediation. Consequently, cognitive diagnostic assessment (CDA) has gained increasing attention among scholars in recent decades (Rupp, Templin, & Henson, 2010; Tu, Cai, & Ding, 2012). CDA refers to diagnostic assessment of individuals' cognitive processes, processing skills, or knowledge structures (collectively termed attributes). As a comprehensive evaluation form combining formative and summative assessment (Zhan, Chen, & Bian, 2016), CDA aims to provide diagnostic feedback reports on individuals' attribute mastery status to teachers or interventionists, thereby facilitating remedial instruction or targeted interventions (Zhan et al., 2018). CDA overcomes the traditional limitation of emphasizing outcomes over processes, aligning with current educational policy orientations such as the *Basic Education Curriculum Reform Outline (Trial)*, which advocates "changing the overemphasis on selection and screening functions in curriculum evaluation, and leveraging evaluation to promote student development, teacher improvement, and instructional practice refinement." Therefore, how to assess scientific literacy within CDA represents a topic with both theoretical and practical significance.

Below, we first provide further interpretation of the PISA 2015 scientific literacy assessment framework to clarify its third-order latent structure. Second, we introduce existing higher-order cognitive diagnosis models (HO-CDM) and explain their limitations. Third, we propose a new multi-order cognitive diagnosis model (MO-CDM) to meet the analytical demands of third-order or higher-order latent structures in CDA and align with the PISA 2015 framework for accurate assessment of scientific literacy. Fourth, we demonstrate the practical applicability of the new model using PISA 2015 science assessment data and interpret the results. Finally, we investigate the parameter recovery of the new model through a simulation study.

## 2. The Third-Order Latent Structure of Scientific Literacy

PISA 2015 conceptualizes scientific competencies as the core of scientific literacy, with the manifestation of scientific competencies requiring sufficient scientific knowledge within specific scientific contexts and being influenced by scientific attitudes. These four dimensions complement each other to constitute scientific literacy, which serves as a higher-order concept encompassing scientific competencies, scientific knowledge, scientific contexts, and scientific attitudes. An individual's level of scientific literacy determines their performance across these four dimensions. According to the *PISA 2015 Assessment and Analytical Framework* (OECD, 2016):

1. Scientific competencies are subdivided into three sub-competencies: explaining phenomena scientifically, evaluating and designing scientific inquiry, and interpreting data and evidence scientifically. Thus, scientific competencies represent a higher-order concept for these three sub-competencies, with individual competency levels determining sub-competency levels.

2. Scientific knowledge is subdivided into three sub-types: content knowledge, procedural knowledge, and epistemic knowledge. Thus, scientific knowledge represents a higher-order concept for these three sub-types, with mastery of scientific knowledge determining mastery of the three sub-types.

3. Scientific contexts are subdivided into three sub-contexts: personal, local/national, and global. Thus, scientific contexts represent a higher-order concept for these three sub-contexts, with familiarity with scientific contexts influencing familiarity with the three sub-contexts.

4. Scientific attitudes are subdivided into three sub-attitudes: interest in science, valuing scientific approaches to inquiry, and environmental awareness. Thus, scientific attitudes represent a higher-order concept for these three sub-attitudes, with overall scientific attitude influencing the three sub-attitudes.

In summary, based on the PISA 2015 scientific literacy assessment framework, scientific literacy comprises a third-order latent structure, as shown in Figure 2. The third-order latent trait is scientific literacy itself, representing the highestorder concept in the PISA 2015 framework. The second-order latent traits include scientific competencies, scientific knowledge, scientific contexts, and scientific attitudes—the four main concepts in the assessment framework. The first-order latent traits consist of 12 specific attributes such as explaining phenomena scientifically and evaluating and designing scientific inquiry.

To assess scientific literacy in CDA, a model capable of analyzing this third-order latent structure is required. Given that no existing CDM can handle third-order latent structures, we must construct a new model to meet this assessment need.

## 3. Multi-Order Cognitive Diagnosis Models

### 3.1 Higher-Order Cognitive Diagnosis Models and Their Limitations

In psychology and education, latent traits may exhibit not only multidimensionality but also hierarchical relationships, termed higher-order latent traits. For example, Figure 2 illustrates the third-order latent structure of scientific literacy, and the Wechsler Adult Intelligence Scale also measures third-order latent traits: the first order includes 13 subtests each measuring a specific trait, the second order groups these into four broader traits (verbal comprehension, perceptual reasoning, working memory, and processing speed), and the third order encompasses these within general intelligence (Ryan & Schnakenberg-Ott, 2003).

The concept of higher-order latent traits is built upon multidimensional latent trait concepts to describe potential structural relationships among multiple latent traits. Based on this, researchers have developed two types of higher-order psychometric models (Chen, Zhan, Wang, Chen, & Cai, 2015): higher-order IRT models built upon multidimensional IRT (de la Torre & Song, 2009; Huang, Wang, Chen, & Su, 2013; Rijmen, Jeon, von Davier, & Rabe-Hesketh, 2014) and higher-order cognitive diagnosis models (HO-CDM) built upon CDMs (de la Torre & Douglas, 2004; Templin, Henson, Templin, & Roussos, 2008; Zhan, Wang, & Li, in press). This study focuses on the latter.

In CDA, recognizing that attribute mastery may be influenced by one or more higher-order latent traits and to reduce the number of parameters to estimate, de la Torre and Douglas (2004) proposed the higher-order latent structural model:

$$\operatorname{logit}(P(\alpha_{nk} = 1|\theta_n)) = \lambda_{0k} + \lambda_{1k}\theta_n \tag{1}$$

where $\operatorname{logit}(P(\alpha_{nk} = 1|\theta_n))$ is the log-odds of examinee $n$ mastering attribute $k$ given the second-order latent trait $\theta_n$; $\lambda_{0k}$ is the difficulty parameter for attribute $k$, and $\lambda_{1k}$ is the discrimination parameter for attribute $k$. The latent structure described in Equation (1) is illustrated in Figure 3. Equation (1) represents the latent structural model, which when combined with a measurement model yields an HO-CDM. For example, combining it with the DINA model (Junker & Sijtsma, 2001; Macready & Dayton, 1977) produces the higher-order

DINA (HO-DINA) model. Due to theoretical limitations of the higher-order latent structural model, the HO-DINA model can only handle data with second-order latent structures and cannot assess the third-order latent structure of scientific literacy, thus failing to meet our research needs.

## 3.2 Multi-Order DINA Model

**3.2.1 Multi-Order Latent Structural Model (MO-LSM)** To address the current lack of CDMs capable of handling third-order or higher-order latent structures, this study draws upon the modeling approach of higher-order IRT models and introduces a linear latent structural model above the existing second-order latent structural model (Equation (1)), proposing a multi-order latent structural model (MO-LSM). First, assuming a multi-order structure for latent traits, let $\theta_{nm}^{(h)}$ represent examinee $n$'s $m$-th latent trait at order $h$ ($h \geq 2$). The linear relationship between $\theta_{nm}^{(h)}$ and higher-order latent traits $\theta_{np}^{(h+1)}$ can be described as:

$$\theta_{nm}^{(h)} = \sum_{p=1}^{P^{(h+1)}} \gamma_{mp}^{(h)} \theta_{np}^{(h+1)} + \varepsilon_{nm}^{(h)} \tag{2}$$

where $\gamma_m^{(h)}$ is the regression vector at order $h$; $\varepsilon_{nm}^{(h)}$ is the residual for the $m$-th latent trait at order $h$ for examinee $n$. Note that beyond linear relationships, Equation (2) could be modified to nonlinear relationships (e.g., polynomial), but given that psychological research typically assumes linear relationships among latent variables (e.g., structural equation modeling) and to reduce model complexity, this study focuses only on linear relationships (de la Torre & Song, 2009; Huang et al., 2013; Rijmen et al., 2014). Introducing Equation (2) into Equation (1) yields the MO-LSM:

$$\text{logit}(P(\alpha_{nk} = 1 | \theta_n^{(2)})) = \lambda_{0k} + \lambda_{1k} \theta_{nm}^{(2)} \tag{3}$$

Assuming conditional independence, MO-LSM posits that lower-order latent traits are independent given higher-order latent traits. Although Equation (3) can theoretically handle multiple orders of latent traits, considering that fourth-order latent traits are rare in practical testing situations and to align with the third-order latent structure of scientific literacy in PISA 2015, this study focuses on third-order latent structural models containing only one third-order latent trait, as shown in Figure 4. This model can be described as:

$$\text{logit}(P(\alpha_{nk} = 1 | \theta_n^{(3)})) = \lambda_{0k} + \lambda_{1k} \left( \sum_{m=1}^{M^{(2)}} \gamma_m^{(2)} \theta_{nm}^{(3)} + \varepsilon_{nm}^{(2)} \right) \tag{4}$$

For model identification, we set $\theta^{(3)} \sim N(0,1)$, which implies $\theta^{(2)} \sim N(0,1)$. Consequently, the correlation between any two second-order latent traits equals

$\gamma_m^{(2)} \times \gamma_{m'}^{(2)}$. When $M^{(2)} = 1$, we have $\varepsilon_{nm}^{(2)} = 0$, and Equation (4) reduces to Equation (1).

**3.2.2 MO-DINA Model** Typically, CDMs consist of two components: a measurement model and a latent structural model (Rupp et al., 2010). The former defines the probability of correct responses, while the latter describes structural relationships among attributes. In Section 3.2.1, we defined MO-LSM. To improve parameter estimation accuracy and efficiency, we selected the Bayesian DINA model incorporating within-item characteristic dependency (Zhan, Jiao, Liao, & Bian, 2018) as the measurement model. Detailed model specifications are provided in the Appendix.

This study employs full Bayesian Markov Chain Monte Carlo (MCMC) algorithms for parameter estimation of the MO-DINA model, implemented using JAGS software (Version 4.3.0). Prior distributions for all parameters are detailed in the Appendix, and corresponding JAGS code is available from the authors. For guidance on implementing Bayesian CDMs using JAGS, see Zhan, Jiao, Man, and Wang (in press).

## 4. Empirical Study: PISA 2015 Science Assessment Data Analysis

### 4.1 Research Questions and Objectives

Through analysis of PISA 2015 science assessment data, this study demonstrates the practical need for and applicability of the MO-DINA model. Based on the third-order latent structure of scientific literacy described above, we aim to assess examinees' performance across all first-order, second-order, and third-order latent traits (attributes). Therefore, this study addresses two questions: (1) Is the MO-DINA model suitable for assessing scientific literacy with a third-order latent structure? If so, (2) Which sub-dimension of scientific literacy has the greatest influence? In other words, what is the core dimension of scientific literacy in PISA 2015?

### 4.2 Participants and Items

**4.2.1 Multi-Order Latent Trait Specification** Based on Section 2, PISA 2015 scientific literacy contains a third-order latent structure. The names of latent traits at each order and their structural relationships are shown in Figure 2. For data analysis, we matched model parameters to multi-order latent traits according to the MO-DINA model: third-order latent trait $\theta^{(3)}$ → scientific literacy; second-order latent traits $\theta_1^{(2)}$ → scientific competencies, $\theta_2^{(2)}$ → scientific knowledge, $\theta_3^{(2)}$ → scientific contexts; first-order attributes: $A_1$ → explaining phenomena scientifically, $A_2$ → evaluating and designing scientific inquiry, $A_3$ → interpreting data and evidence scientifically, $A_4$ → content knowledge, $A_5$ → procedural knowledge, $A_6$ → epistemic knowledge, $A_7$ → personal context, $A_8$

→ local/national context, $A_9$ → global context. Note that in the second-order latent traits, scientific attitudes were obtained through student questionnaires rather than cognitive item data, so this dimension is not included in the current study.

**4.2.2 Data Sources and Processing** According to the *PISA 2015 Technical Report* (OECD, 2017) "Appendix A: Item Pool Classification," the data cleaning process was as follows: (1) Selected 18 items from the "2015 field trial and main survey cluster" S01, totaling 47,548 examinees; (2) Selected the China (QCH) sample, totaling 1,079 examinees; (3) Recoded "not reached" and "no response" as missing values (NA); (4) Deleted 3 examinees with missing responses on all 18 items, retaining 1,076 examinees; (5) Treated all remaining missing values as missing completely at random. Full Bayesian MCMC algorithms can compute posterior distributions for missing values based on other parameter estimates, an "automatic imputation" process requiring no additional specifications. Additionally, item DS519Q01 was originally polytomously scored ($Y_{ni} \in \{0, 1, 2\}$). Since the MO-DINA model currently only handles dichotomous items, we dichotomized this item's scores: $0\$\to 0, 1\to 0, 2\to\$1$. The final cleaned dataset contained dichotomous response data from $N = 1,076$ examinees on $I = 18$ items. The Q-matrix mapping between attributes and items is shown in Table 1 .

Three models were fitted and compared: MO-DINA, HO-DINA, and DINA. For the MO-DINA model, the multi-order latent structure was specified according to Figure 2 (excluding scientific attitudes). For the HO-DINA model, we assumed first-order attributes were directly influenced by scientific literacy, ignoring second-order latent traits by constraining $\gamma_m^{(2)} = 1$. For the DINA model, all multi-order latent structures were ignored, using an unstructured latent structural model.

All three models used two Markov chains (random starting points), each with 10,000 iterations, including 5,000 burn-in iterations and thinning of 1, retaining 10,000 iterations for parameter estimation. Convergence was assessed using the Potential Scale Reduction Factor (PSRF) (Brooks & Gelman, 1998). All parameters in this study achieved PSRF $< 1.2$, indicating convergence.

Model-data fit was evaluated using AIC, BIC, and DIC, where smaller values indicate better relative fit. Additionally, posterior predictive model checking (PPMC) was used to assess absolute fit, with posterior predictive probability (ppp) values near 0.5 indicating adequate fit and values below 0.05 or above 0.95 indicating misfit.

Table 2 presents model-data fit indices for the three models. First, based on ppp values, all three models showed adequate fit to the data. Second, according to the four relative fit indices, the DINA model showed the poorest fit, suggesting that higher-order latent structures should be considered for this dataset. Third, among the four relative fit indices, –2LL and AIC indicated better relative fit for the MO-DINA model, while BIC and DIC favored the HO-DINA model, due to BIC and DIC's stronger penalties for model complexity. Since the HO-DINA model is a special case of the MO-DINA model (i.e., constrained $\gamma_m^{(2)} = 1$), a likelihood ratio test ($\Delta$–2LL = 13, $df = 3$, $p < 0.05$) indicated significant differences between the models, favoring the MO-DINA model. Finally, considering the research questions and objectives, we concluded that the MO-DINA model is most appropriate for this study. Subsequent interpretations are based on MO-DINA model results.

Table 3 presents item parameter estimates. Overall, the quality of the 18 items was modest, with some items showing guessing or slipping parameters around 0.8. This is also evident from item discrimination indices ($IDI_i = 1 - s_i - g_i$) (de la Torre, 2008), with some items falling below 0.2. Possible reasons include: (1) incomplete Q-matrix specification (Köhn & Chiu, 2017); (2) items tapping attributes beyond those specified in the Q-matrix. Table 4 presents the mean vector and variance-covariance matrix of logit-transformed item parameters, showing high negative correlations between the two types of item parameters, consistent with Zhan et al. (2018).

Regarding higher-order latent trait estimates, first, the distributions of the one third-order and three second-order latent trait estimates were generally similar due to high correlations among them (three regression coefficients: 0.847 (SE = 0.094), 0.973 (SE = 0.025), and 0.927 (SE = 0.057), yielding intercorrelations of approximately 0.8). Note that high statistical correlations among traits do not necessarily indicate they are the same trait. For example, although height and weight are highly correlated, they are distinct traits. Therefore, when traits are highly correlated, whether a general higher-order trait can encompass them requires further theoretical justification. Based on the PISA 2015 framework, we argue that these three second-order latent traits differ in definition and connotation and should not be treated as identical. Additionally, we analyzed the same data using the HO-DINA model and a unidimensional two-parameter logistic model (Birnbaum, 1968), finding that the higher-order latent trait estimates from the HO-DINA model correlated at 0.996, and correlated at 0.936 with the latent trait estimates from the unidimensional model, indicating high convergence in estimating "scientific literacy" while demonstrating that the MO-DINA model provides more analytical information.

Figure 5 [Figure 5: see original paper] presents estimates of higher-order structural parameters, including regression coefficients between third-order and second-order latent traits and attribute discrimination parameters between second-order latent traits and attributes. First, all three regression coefficients approached 1, supporting PISA 2015's treatment of competencies, knowledge, and contexts as primary components of scientific literacy. Second, based on the magnitude of these coefficients: knowledge has the greatest influence on scientific literacy, followed by contexts, with competencies having the smallest effect. Third, based on attribute discrimination parameters: (1) explaining phenomena scientifically has the greatest influence on competencies; (2) procedural knowledge has the greatest influence on knowledge; (3) local/national contexts have the greatest influence on contexts.

Table 5 presents examples of individual diagnostic results. Using the MO-DINA model, we obtained diagnostic classification results for nine attributes as well as estimates on multi-order latent traits. For example, examinees 2 and 23 showed identical attribute patterns but differed in their multi-order latent trait estimates, indicating differential probabilities of attribute mastery.

Overall, analysis of PISA 2015 science assessment data suggests that the MO-DINA model meets our analytical needs, achieving objective assessment of scientific literacy while aligning with the PISA 2015 assessment framework.

## 5. Simulation Study: Parameter Recovery Investigation

### 5.1 Design and Analysis

After demonstrating the practical applicability of the MO-DINA model, we conducted a simulation study to examine its parameter recovery. Simulation conditions were partially based on the empirical analysis results above, using the third-order latent structure shown in Figure 7 [Figure 7: see original paper]: one third-order latent trait, three second-order latent traits, and $K = 9$ attributes; $I = 30$ items; Q-matrix specification shown in Figure 6 [Figure 6: see original paper]. Item parameters were generated as follows: $(\text{logit}(g_i), \text{logit}(s_i))' \sim N(\mu, \Sigma)$, where $\mu_\beta = \mu_\delta = -2.197$, $\Sigma = [1, -0.6; -0.6, 1]$; attribute intercept vector $\lambda_0 = (-1, 0, 1, -1, 0, 1, -1, 0, 1)$; all attribute discriminations set to $\lambda_{1mk} = 1.5$, assuming moderate correlations among attributes; sample size $N = 1,000$; third-order latent traits generated from standard normal distribution; three loadings between third-order and second-order latent traits set to $\gamma_m^{(2)} = 0.8$, assuming correlations of 0.64 among second-order latent traits. Iteration numbers, burn-in, and other settings matched those in the empirical study, with all parameters achieving PSRF $< 1.2$, indicating convergence. Bias, root mean square error (RMSE), and Pearson correlation coefficients (Cor) were used to examine recovery of continuous variables (e.g., item parameters, latent traits). Attribute correct classification rate (ACCR) and pattern correct classification rate (PCCR) were used to examine recovery of attributes.

Figure 7 [Figure 7: see original paper] presents item parameter recovery. In terms of bias, absolute bias for most item parameters was less than 0.01, with mean absolute bias for guessing and slipping parameters of 0.002 and 0.004, respectively. In terms of RMSE, all item parameter RMSEs were less than 0.05, with mean RMSEs for guessing and slipping parameters of 0.018 and 0.026, respectively. Notably, RMSE for guessing parameters decreased with increasing numbers of attributes measured by items, while RMSE for slipping parameters increased with increasing numbers of attributes, consistent with previous research (e.g., de la Torre, 2009; Zhan, Jiao, Liao, et al., 2018). Correlations between estimated and true item parameters were high, with Cor for guessing and slipping parameters of 0.981 and 0.964, respectively. Overall, item parameter recovery for the MO-DINA model was satisfactory.

Figure 8 [Figure 8: see original paper] presents ACCR for attribute parameters. ACCR for all nine attributes exceeded 0.900, indicating excellent recovery of individual attributes. PCCR was 0.512. Considering that nine attributes yield $2^9 = 512$ possible attribute patterns, this classification rate meets expectations based on previous research.

Table 6 presents recovery of higher-order latent trait parameters. First, recovery was similar across the four higher-order latent traits, with mean absolute bias of approximately 0.1, mean RMSE of approximately 0.69, and Cor exceeding 0.7 for 1,000 examinees. These results are comparable to previous HO-DINA model studies (e.g., de la Torre & Douglas, 2004; de la Torre, 2009; Zhan et al., 2018). Overall, higher-order latent trait parameter recovery was adequate for practical applications.

Table 7 presents recovery of higher-order structural parameters. First, recovery of attribute difficulty parameters was better than that of attribute discrimination parameters, consistent with previous HO-DINA research. Second, recovery of regression coefficients between third-order and second-order latent traits was also satisfactory, with RMSEs less than 0.08. Overall, higher-order structural parameter recovery was good.

## 6. Summary and Discussion

To achieve objective and accurate assessment of scientific literacy, this study first proposed a third-order latent structure of scientific literacy based on the PISA 2015 assessment framework. Given the absence of CDMs capable of handling third-order latent structures, we proposed a multi-order cognitive diagnosis modeling approach and, using the DINA model as an example, constructed the multi-order DINA (MO-DINA) model. The new model employs full Bayesian MCMC algorithms for parameter estimation and aligns with the PISA 2015 framework to meet the need for objective and accurate assessment of scientific literacy. We demonstrated the practical applicability of the new model using PISA 2015 science assessment data and investigated its parameter recovery through simulation studies. Empirical results showed that the MO-DINA model works well in real data analysis, and simulation results indicated that the proposed Bayesian MCMC algorithm provides accurate parameter estimation.

Although the MO-DINA model was developed specifically for the third-order latent structure of scientific literacy in PISA 2015, and because it extends the HO-DINA model, it can theoretically be applied to other assessments containing second-order and higher-order latent structures, such as the Trends in International Mathematics and Science Study (TIMSS) and China's National Compulsory Education Quality Monitoring. However, this study does not claim that all assessments with multi-order latent structures or all assessments of scientific literacy must use the MO-DINA model. Rather, from the perspective of "assessment for learning," we provide a new analytical perspective and method to enrich the available options for data analysis. In practice, model selection can be based on test development theory, actual assessment needs, or data-driven approaches using model-data fit indices (e.g., AIC, BIC, DIC) to obtain objective, accurate, and need-satisfying results.

It should be emphasized that higher-order models are generally considered when there are three or more lower-order latent traits. For example, for the second-order LSM (see Equation (1)), when $K = 3$, an unstructured latent structural model requires estimating $2^3 - 1 = 7$ structural parameters, while the second-order LSM requires only 6 parameters (including 3 attribute discriminations and 3 attribute difficulties). For the relationship between third-order and second-order latent traits, when there are 3 second-order latent traits, both direct estimation of correlations among them and estimation of loadings from third-order to second-order traits require 3 parameters. When the number of second-order latent traits exceeds 3, higher-order structures can reduce the number of parameters to estimate. For instance, for the third-order latent structure in Figure 5, direct use of the DINA model would require estimating $2^9 - 1 = 511$ structural parameters, while the MO-DINA model requires only 21 structural parameters (including 9 attribute discriminations, 9 attribute difficulties, and 3 loadings), substantially reducing parameters. However, a three-dimensional second-order DINA model would also require only 21 structural parameters (including 9 attribute discriminations, 9 attribute difficulties, and 3 correlations) but could not measure the "scientific literacy" dimension. Therefore, the decision to use higher-order models can be considered from both theoretical (assessment framework) and parsimony perspectives, but the ultimate justification must be theoretical, as not all latent traits are suitable for higher-order structures. For example, the Big Five personality dimensions should not be explained by a higher-order "personality" trait, as the five dimensions are theoretically independent (though empirical analyses may show low correlations).

Although this study divided scientific literacy into a third-order latent structure, the first-order attributes remain relatively coarse-grained, whereas CDA is typically more suitable for assessing fine-grained attributes (see Leighton & Gierl, 2007; Zhan et al., 2016). In fact, based on the PISA 2015 framework, the first-order attributes in this study could be further subdivided into finer-grained concepts. For example, $A_1$ "explaining phenomena scientifically" could be further divided into "recall and apply appropriate scientific knowledge" and "offer explanatory hypotheses" (see OECD, 2016, Table 2.4a). Although a fourth-order MHO-DINA model could theoretically be used for further analysis, the *PISA 2015 Technical Report* does not provide the specific mapping between items and fine-grained concepts (i.e., no corresponding Q-matrix), so this study focused only on assessing the third-order latent structure of scientific literacy. If needed, future research could apply third-order IRT models (e.g., Huang et al., 2013) to analyze this data and compare results with those from this study.

Several limitations warrant further investigation: (1) Although focusing on the latent structural model, this study only used the DINA model as the measurement model; performance with other measurement models should be explored. (2) Attribute hierarchy structures (Leighton, Gierl, & Hunka, 2004) were not considered; integrating hierarchical structures into multi-order latent structures deserves attention (e.g., Zhan, Ma, Jiao & Ding, in press). (3) Only dichotomous attributes were considered, not more refined polytomous attributes (Karelitz, 2004); extending MO-LSM to polytomous attributes is meaningful (e.g., Zhan, Wang et al., in press). (4) The multi-order latent structure was assumed correctly specified, but real-world specification errors warrant investigation of MO-DINA model performance under misspecification. (5) MO-DINA only considered a single data source (item responses), not process data such as response times or click sequences; integrating process data into the modeling framework is valuable (e.g., Liu, Liu, & Li, 2018; Zhan et al., 2018). (6) MO-DINA only handles cross-sectional data, not longitudinal data; further extension is needed (e.g., Li, Cohen, Bottge, & Templin, 2016; Zhan, Jiao, Liao & Li, in press). (7) The empirical analysis did not include the scientific attitudes dimension; integrating attitudinal data from student questionnaires with cognitive dimensions from item responses deserves future exploration.

## References

Birnbaum, A. (1968). Some latent trait models and their use in inferring a student's ability. In F. M. Lord & M. R. Novick (Eds.). *Statistical theories of mental test scores*. Addison-Wesley, Reading, MA.

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4), 434–455.

Chen, F., Zhan, P., Wang, L., Chen, C., & Cai, M. (2015). The development and application of higher-order item response models. *Advances in Psychological Science*, 23, 150–157.

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(4), 343–362.

de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, 34(1), 115–130.

de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, 69(3), 333–353.

de la Torre, J., & Song, H. (2009). Simultaneously estimation of overall and domain abilities: A higher-order IRT model approach. *Applied Psychological Measurement*, 33(8), 620–639.

Gao, H. B. (2011). Results of the eighth survey on Chinese citizens' scientific literacy were released. *Bulletin of National Natural Science Foundation of China*, 25, 63–64.

Hu, Y., Yang, S., & Lu, K. (2012). The research of assessment tools of adolescents' scientific literacy and its quality analysis. *Education Research Monthly*, 3, 16–21.

Huang, H.-Y., Wang, W.-C., Chen, P.-H., & Su, C.-M. (2013). Higher-order item response models for hierarchical latent traits. *Applied Psychological Measurement*, 37(8), 619–637.

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.

Karelitz, T. M. (2004). *Ordered category attribute coding framework for cognitive assessments* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.

Kang, C., & Xin, T. (2010). New development in test theory: Multidimensional item response theory. *Advances in Psychological Science*, 18(3), 530–536.

Köhn, H.-F., & Chiu, C.-Y. (2017). A procedure for assessing the completeness of the Q-matrices of cognitively diagnostic tests. *Psychometrika*, 82(1), 112–132.

Leighton, J., & Gierl, M. (Eds.). (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge University Press.

Leighton, J. P., Gierl, M. J., & Hunka, S. M. (2004). The attribute hierarchy method for cognitive assessment: A variation on Tatsuoka's rule-space approach. *Journal of Educational Measurement*, 41(1), 205–237.

Li, F., Cohen, A., Bottge, B, & Templin, J. (2016). A latent transition analysis model for assessing change in cognitive skills. *Educational and Psychological Measurement*, 76(2), 181–204.

Liu, H., Liu, Y., & Li, M. (2018). Analysis of process data of PISA 2012 computer-based problem solving: Application of the modified multilevel mixture IRT model. *Frontiers in Psychology*, 9, 1372.

Liu, K., & Li, C. (2015). The content and characteristic of PISA 2015 draft science framework. *Comparative Education Review*, 37(7), 98–105.

Macready, G. B., & Dayton, C. M. (1977). The use of probabilistic models in the assessment of mastery. *Journal of Educational and Behavioral Statistics*, 2(2), 99–120.

Miller, J. D. (1983). Scientific literacy: A conceptual and empirical review. *Daedalus*, 112(2), 29–48.

OECD. (2006). *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006*. Paris: PISA, OECD Publishing.

OECD. (2016). *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic and Financial Literacy.* Paris: PISA, OECD Publishing.

OECD. (2017). *PISA 2015 Technical Report.* Paris: PISA, OECD Publishing.

Qin, H., & Qian, Y. (2008). A survey report on Shanghai adolescents' scientific literacy. *Research in Educational Development*, (24), 31–35.

Ren, L., Zhang, C., & He, W. (2013). Constructing and analysis of the model of how the factors affect the scientific literacy of Chinese citizens and a comparative investigation. *Studies in Science of Science*, 31(7), 983–990.

Reckase, M. D. (2009). *Multidimensional item response theory.* New York: Springer.

Rijmen, F., Jeon, M., von Davier, M., & Rabe-Hesketh, S. (2014). A third-order item response theory model for modeling the effects of domains and subdomains in large-scale educational assessment surveys. *Journal of Educational and Behavioral Statistics*, 39(4), 235–256.

Roos, J. M. (2014). Measuring science or religion? A measurement analysis of the National Science Foundation sponsored science literacy scale 2006–2010. *Public Understanding of Science*, 23(7), 797–813.

Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic Measurement: Theory, Methods, and Applications.* New York, NY: Guilford Press.

Ryan, J. J., & Schnakenberg-Ott, S. D. (2003). Scoring reliability on the Wechsler Adult Intelligence Scale-Third Edition (WAIS-III). *Assessment*, 10(2), 151–159.

Templin, J. L., Henson, R. A., Templin, S. E., & Roussos, L. (2008). Robustness of hierarchical modeling of skill association in cognitive diagnosis models. *Applied Psychological Measurement*, 32(7), 559–574.

The Ministry of Education of the People's Republic of China. (2017). *Compulsory education primary school curriculum standards.* http://www.moe.edu.cn/srcsite/A26/s8001/201702/t2017

Tu, D., Cai, Y., & Ding, S. (2012). *Cognitive diagnosis: Theory, Methods, and Applications.* Beijing: Beijing Normal University Publishing Group.

Wang, W.-C., & Chen, P.-H. (2004). Implementation and measurement efficiency of multidimensional computerized adaptive testing. *Applied Psychological Measurement*, 28(5), 295–316.

Zhan, P., Chen, P., & Bian, Y. (2016). Using confirmatory compensatory multidimensional IRT models to do cognitive diagnosis. *Acta Psychologica Sinica*, 48(10), 1347–1356.

Zhan, P., Jiao, H., & Liao, D. (2018). Cognitive diagnosis modelling incorporating item response times. *British Journal of Mathematical and Statistical Psychology*, 71(2), 218–240.

Zhan, P., Jiao, H., Liao, D., & Li, F. (in press). A longitudinal higher-order diagnostic classification model. *Journal of Educational and Behavioral Statistics.*

Zhan, P., Jiao, H., Liao, M., & Bian, Y. (2018). Bayesian DINA modeling incorporating within-item characteristic dependency. *Applied Psychological Measurement.* Advanced online publication. URL https://doi.org/10.1177/0146621618781594

Zhan, P., Jiao, H., Man, K., & Wang, L. (in press). Using JAGS for Bayesian cognitive diagnosis modeling: A tutorial. *Journal of Educational and Behavioral Statistics.*

Zhan, P., Ma, W., Jiao, H., & Ding, S. (in press). A sequential higher-order latent structural model for hierarchical attributes in cognitive diagnostic assessments. *Applied Psychological Measurement.*

Zhan, P., Wang, W.-C., & Li, X. (in press). A partial mastery, higher-order latent structural model for polytomous attributes in cognitive diagnostic assessments. *Journal of Classification.*

Zhan, P., Wang, W.-C., & Wang, L. (2013). Testlet response theory: an introduction and new developments. *Advances in Psychological Science*, 21(12), 2265–2280.

---

# Appendix

## 1. MO-DINA Model

The measurement model employs the Bayesian DINA model incorporating within-item characteristic dependency (Zhan, Jiao, Liao, & Bian, 2018), expressed as:

$$P(Y_{ni} = 1|\alpha_n, \Psi_i) = (1 - s_i)^{\eta_{ni}} g_i^{1-\eta_{ni}} \tag{A1}$$

where $Y_{ni}$ is examinee $n$'s response to item $i$; $\Psi_i = (\beta_i, \delta_i)'$ is the vector of item parameters on the logit scale (typically negatively correlated), related to conventional DINA guessing and slipping parameters as $\text{logit}(g_i) = \beta_i$ and $\text{logit}(s_i) = \delta_i$; $q_{ik}$ are Q-matrix elements where $q_{ik} = 1$ indicates item $i$ measures attribute $k$ and $q_{ik} = 0$ otherwise; and $\eta_{ni} = \prod_{k=1}^{K} \alpha_{nk}^{q_{ik}}$ is the ideal response.

Combining this measurement model with Equation (4) from the main text yields the MO-DINA model.

## 2. Prior Distribution Specifications for MO-DINA Model Parameters

First, based on local independence, $Y_{ni} \sim \text{Bernoulli}(P(Y_{ni} = 1|\alpha_n, \Psi_i))$, and $\alpha_{nk}|\theta_n^{(2)} \sim \text{Bernoulli}(P(\alpha_{nk} = 1|\theta_n^{(2)}))$.

Second, for item parameters, following Zhan, Jiao, Liao, et al. (2018), we specify:

$$\begin{pmatrix}\text{logit}(g_i)\\ \text{logit}(s_i)\end{pmatrix} \sim N(\mu, \Sigma) \tag{A2}$$

where $\mu = (\mu_\beta, \mu_\delta)'$ is the mean of logit-transformed item parameters and $\Sigma$ is the variance-covariance matrix with $\Sigma = \begin{pmatrix}\sigma_\beta^2 & \rho_{\beta\delta}\sigma_\beta\sigma_\delta\\ \rho_{\beta\delta}\sigma_\beta\sigma_\delta & \sigma_\delta^2\end{pmatrix}$, where $\rho_{\beta\delta}$ is the correlation between logit-transformed parameters. Hyper-priors were set as $\mu_\beta \sim N(-1.096, 4)$ and $\mu_\delta \sim N(-1.096, 4)$. Since $\text{logit}(-1.096) \approx 0.25$, this aligns with theoretical guessing probabilities for four-option multiple-choice items. Additionally, $\Sigma \sim \text{InvWishart}(R, 2)$, where $R$ is a 2$×$2 identity matrix.

Third, for higher-order latent traits, following Huang et al. (2013), we specify:

$$\theta_n^{(3)} \sim N(0,1), \quad \varepsilon_{nm}^{(2)} \sim N(0, 1-\gamma_m^{(2)2}), \quad \gamma_m^{(2)} \sim N(0.5, 0.25)I(-1,1) \tag{A3}$$

Finally, for higher-order structural parameters, following Zhan, Jiao, and Liao (2018), we specify:

$$\lambda_{0k} \sim N(0,4), \quad \lambda_{1mk} \sim N(0,4)I(\lambda_{1mk} > 0) \tag{A4}$$

*Note: Figure translations are in progress. See original paper for figures.*

*Source: ChinaXiv — Machine translation. Verify with original.*