Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 28 Number 11, August 2023 ISSN 1531-7714

Using a Bayesian Estimation to Examine Attribute Hierarchies of the 2007 TIMSS Mathematics Test: A Demonstration Using R Packages[1]

Chia-Ling Hsu, *Hong Kong Examinations and Assessment Authority*
Yi-Hsin Chen, *University of South Florida*
Yi-Jhen Wu, *Technische Universität Dortmund*

Correct specifications of hierarchical attribute structures in analyses using diagnostic classification models (DCMs) are pivotal because misspecifications can lead to biased parameter estimations and inaccurate classification profiles. This research is aimed to demonstrate DCM analyses with various hierarchical attribute structures via Bayesian estimation using freely available R packages, including *CDM* and *R2jags*. We illustrated a step-by-step procedure in R with an eighth-grade mathematics test from the 2007 Trends in International Mathematics and Science Study (TIMSS).

Keywords: hierarchical attribute structure, cognitive diagnosis, Bayesian estimation, LCDM, TIMSS

Introduction

Diagnostic classification models (DCMs; Rupp et al., 2010) have received considerable attention in the field of education over the past two decades. DCMs are used to evaluate a respondent's mastery status on a set of fine-grained discrete *latent skills* (also known as *attributes*) to derive tailored information on his/her learning. Such a personalized diagnosis yields more insights on students' learning, which in turn can be used by teachers to customize instruction and accordingly optimize learning outcomes (e.g., Birenbaum et al., 2004; Chen, 2012; Dogan & Tatsuoka, 2008; Lee et al., 2011; Tatsuoka et al., 2004).

The DCM literature indicates that misspecifications of skill structures lead to detrimental effects on the recovery of parameters and the accuracy of examinee classifications (e.g., Liu, 2018; Liu et al., 2017; Templin & Bradshaw, 2014; Templin et al., 2008). In other words, DCM analyses with misspecifications of attribute structures are expected to result in inaccurate examinations of model and item fit and the classification mastery statuses of examinees. For DCM parameter estimation, maximum likelihood estimation (MLE) and Bayesian estimation are two most widely used methods. Yet, DCM analyses with MLE have been suggested to be insensitive to capture differences across various attribute structures. Liu's study (2018), for example, showed that the differences of fit results between the true and misspecified hierarchical attribute structures were very small. Thus, this study is aimed at providing a DCM tutorial that uses Bayesian estimation with R packages for researchers and practitioners. Because Bayesian computation is more complex than MLE, most statistical software does not provide Bayesian results.

[1] The authors appreciate the Editors' and the anonymous reviewers' valuable remarks on this article. Furthermore, the authors would like to thank Juntao Wang and Xue Zhang for graciously contributing the TIMSS 2007 data for the instructional presentation.

For example, The R package *CDM* for DCM analyses can only yield MLE results. In other words, if researchers would like to conduct DCM analyses with Bayesian estimation, they have to utilize specialized software (e.g., JAGS; Plummer, 2003). Meanwhile, the analytical codes for Bayesian estimation may not be transparent. Thus, we will illustrate the procedures using the open-source software R in this paper (R Core Team, 2012), so that researchers may refer to the procedures and produce the analytical codes on their own.

Hence, the present paper was intended to describe how two packages, *CDM* (Robitzsch et al., 2019) and *R2jags* (Su & Yajima, 2015), can be combined to conduct a DCM analysis with different attribute hierarchies by means of Bayesian estimation. Given that the R software and the concept of Bayesian estimation are complicated for practitioners, this study is highly beneficial to researchers and practitioners in education and psychology. Before providing a step-by-step walk-through of the analysis, we begin with a short overview of DCMs.

Diagnostic Classification Models

There are two essential prerequisites for conducting DCM analyses: one is a Q-matrix, which is an item-to-attribute matrix that represents the relationship between items and attributes, and the other is an attribute profile, which is intended to portray each examinee's mastery status. The Q-matrix is a $J \times K$ matrix, where J (j = 1, 2, …, J) represents items, and K (k = 1, 2, …, K) denotes attributes. The Q-matrix is specified by content experts to identify which attributes are measured by which items (Tatsuoka, 1983). In the Q-matrix, entry q_{jk} is 1 if the correct response to item j requires attribute k, and it is 0 otherwise. The attribute profile is defined as $\alpha_i = (\alpha_{i1}, \alpha_{i2}, \dots \alpha_{iK})'$ to represent the mastery status of examinee i; $\alpha_{ik} = 1$ indicates that examinee i has mastered attribute k, and $\alpha_{ik} = 0$ indicates otherwise.

DCMs can generally be divided into compensatory, non-compensatory, and general models. In compensatory models, a lack of non-mastery in one attribute can be compensated by mastery of another attribute; examples of these models are the *deterministic input noisy output "OR" gate* (DINO) model (Templin & Henson, 2006) and the *compensatory reparametrized unified model* (Hartz, 2002). In non-compensatory models, the compensatory feature is infeasible, and thus, all required attributes are necessary to solve an item; examples of these models include the *deterministic input noisy output "AND" gate* (DINA) model (Junker & Sijtsma, 2001) and the *non-compensatory reparametrized unified model* (NC-RUM) (DiBello et al., 1995; Hartz, 2002). General models allow for both compensatory and non-compensatory relationships within a single test. Such flexibility renders general models more practical for real data analysis than their compensatory and non-compensatory counterparts. General DCMs include the *generalized DINA model* (de la Torre, 2011), the *general diagnostic model* (von Davier, 2005), and the *log-linear cognitive diagnostic model* (LCDM) (Henson et al., 2009). In this study, we used the LCDM to demonstrate attribute hierarchy analyses in R for two reasons: (1) The LCDM encompasses many popular DCMs (e.g., the DINA, DINO, and NC-RUM models), and (2) it offers flexibility in modeling attribute structures by setting redundant parameters to 0 when different attribute hierarchies are represented (Henson et al., 2009; Templin & Bradshaw, 2014).

Structures of Attribute Hierarchy

A general assumption in the learning process is that if a person is to master high-level skills, he/she should first attain proficiency in low-level skills (Darling-Hammond et al., 2015; Entwistle & Ramsden, 2015). In geometry learning, for instance, before a student can analyze the features of a square, he/she has to learn the skill to recognize this shape. On this basis, the acquisition of attributes is a hierarchical learning process, with the mastery of certain attributes being a prerequisite to the mastery of others.

Leighton et al. (2004) introduced four types of attribute hierarchy: *linear*, *convergent*, *divergent*, and *unstructured hierarchies* (Figure 1). Let us consider the case of a *linear hierarchy* with five attributes that are sequentially ordered; that is, Attribute 1 is a prerequisite to Attribute 2, Attributes 1 and 2 are prerequisites to Attribute 3, and so on. Accordingly, an examinee who has mastered Attribute 5 is anticipated to have mastered Attributes 1 to 4. Put differently, mastery of a higher-order attribute assumes mastery of all lower-order attributes. In a *convergent hierarchy*, different branches converge to a common attribute. As shown in Figure 1, two paths can be traced from

Attributes 1 to 5, wherein an examinee who has become proficient in Attribute 5 is expected to have acquired one or more of the preceding attributes (e.g., Attributes 1–3 or Attributes 1, 2, and 4). Similar to a *linear hierarchy*, this kind of hierarchy ends at a single point. In a *divergent hierarchy*, multiple branches diverge from a common attribute. Hence, an examinee who has possessed Attribute 5 is anticipated to have thoroughly grasped all the preceding attributes on a specific path (e.g., Attributes 1 and 3). In an *unstructured hierarchy*, Attribute 1 is a prerequisite to Attributes 2 to 5, but the relationship among Attributes 2 to 5 is unknown. Correspondingly, an examinee who possesses Attribute 5 is anticipated to have mastered only Attribute 1.

LCDM with Different Hierarchical Attribute Structures

Equation (1) shows how parameters are established to represent different structures of attribute hierarchy. In the LCDM, the probability of answering item j correctly under an attribute profile $\alpha_i = (\alpha_{i1}, \alpha_{i2}, \ldots \alpha_{iK})'$ for examinee i is as follows:

$$P\left(X_{ij} = 1 \middle| \boldsymbol{\alpha_i}\right) = \frac{\exp\left[\lambda_{j,0} + \boldsymbol{\lambda}_j^T \mathbf{h}(\boldsymbol{\alpha}_i, \mathbf{q}_j)\right]}{1 + \exp\left[\lambda_{j,0} + \boldsymbol{\lambda}_j^T \mathbf{h}(\boldsymbol{\alpha}_i, \mathbf{q}_j)\right]} \qquad (1)$$

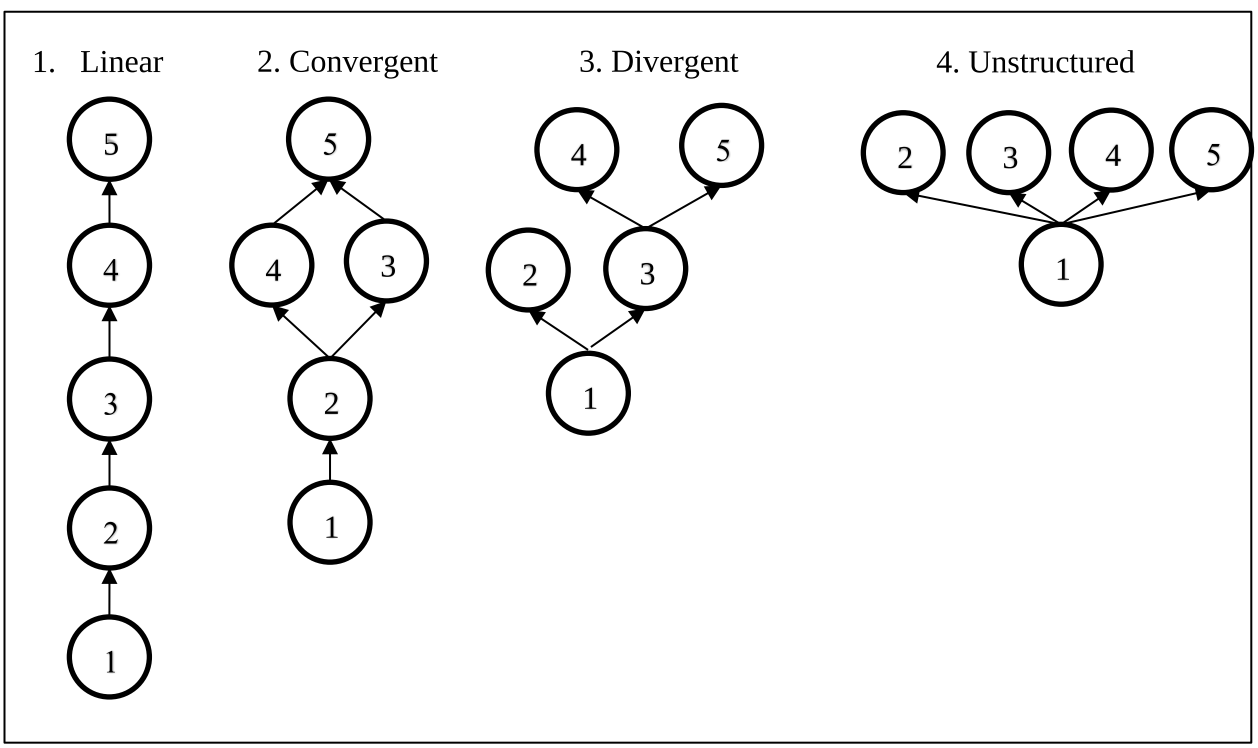
where X_{ij} is the response of examinee i to item j, $\lambda_{j,0}$ denotes the intercept parameter that represents the log-odds of a correct response from an examinee who has not mastered any of the attributes under item j, $\boldsymbol{\lambda}_j^T$ represents a vector of main and interaction effect parameters for item j, and $\mathbf{h}(\boldsymbol{\alpha}_i, \mathbf{q}_j)$ is a mapping function that describes how latent attribute profile $\boldsymbol{\alpha}_i$ is combined with the main and interaction effect parameters in the model.

Let us take two attributes as examples. The assumption is that an item measures two attributes, resulting in entries q_{j1} and q_{j2} being equal to 1 in the Q-matrix. On the basis of Equation (1), therefore, the probability of correctly answering item j is expressed as

$$P\left(X_{ij} = 1 \middle| \alpha_i\right) = \frac{\exp\left(\lambda_{j,0} + \lambda_{j,1,(1)}\alpha_{i1} + \lambda_{j,1(2)}\alpha_{i2} + \lambda_{j,2,(1,2)}\alpha_{i1}\alpha_{i2}\right.}{1 + \exp\left(\lambda_{j,0} + \lambda_{j,1,(1)}\alpha_{i1} + \lambda_{j,1(2)}\alpha_{i2} + \lambda_{j,2,(1,2)}\alpha_{i1}\alpha_{i2}\right.} \qquad (2)$$

where $\lambda_{j,1,(1)}$ and $\lambda_{j,1,(2)}$ are the parameters for the two main effects associated with a_1 and a_2, respectively, and $\lambda_{j,2,(1,2)}$ is the parameter for the two-way

Figure 1. Four types of hierarchical attribute structures featuring five attributes



interaction effect between a_1 and a_2. If mastering a_1 is a prerequisite for mastering a_2, a *linear* attribute hierarchy can be used to model such a sequential structure, with the main effect directly eliminated for nested attribute a_2. Therefore, the item response function can be re-written thus:

$$P(X_{ij} = 1|\alpha_i) = \frac{\exp\,[\lambda_{j,0}+\lambda_{j,1,(1)}\alpha_{i1}+\lambda_{j,2,(2(1))}\alpha_{i1}\alpha_{i2}]}{1+\exp\,[\lambda_{j,0}+\lambda_{j,1,(1)}\alpha_{i1}+\lambda_{j,2,(2(1))}\alpha_{i1}\alpha_{i2}]} \quad (3)$$

where $\lambda_{j,2,(2(1))}$ is the parameter for the two-way interaction effect between a_1 and a_2 when a_1 is a prerequisite of a_2. Likewise, if mastering a_2 is a prerequisite for mastering a_1, the parameter for the main effect of a_2 ($\lambda_{j,1,(2)}$) is retained. Then, Equation (2) can be re-written into

$$(X_{ij} = 1|\alpha_i) = \frac{\exp[\lambda_{j,0}+\lambda_{j,1,(2)}\alpha_{i2}+\lambda_{j,2,(1(2))}\alpha_{i1}\alpha_{i2}]}{1+\exp[\lambda_{j,0}+\lambda_{j,1,(2)}\alpha_{i2}+\lambda_{j,2,(1(2))}\alpha_{i1}\alpha_{i2}]} \quad (4)$$

where $\lambda_{j,2,(1(2))}$ is the parameter for the two-way interaction effect between a_1 and a_2 when a_2 is a prerequisite of a_1.

A Demonstration with a Mathematics Test

A dataset of American eighth-grade mathematics achievement was extracted from the 2007 Trends in International Mathematics and Science Study (TIMSS). The mathematics test items were taken from Booklet 1, which measures three attributes (Lee et al., 2013; Ma, 2019; Zhang & Wang, 2020): (a) whole numbers and integers (Attribute 1); (b) fractions, decimals, and percentages (Attribute 2); and (c) data analysis and probability (Attribute 3). From Zhang and Wang's (2020) dataset, we adopted 12 items that were dichotomously scored and to which 544 students responded in our work. Table 1 presents the Q-matrix that describes the relationships between the three attributes and the 12 items. In the matrix, each item measures only one attribute; that is, three items measure Attribute 1, six items measure Attribute 2, and three items measure Attribute 3.

To demonstrate how implementing DCM analyses with different structures of attribute hierarchy in R via Bayesian estimation, we elaborate on the following steps: (1) installing/loading R packages, (2) retrieving datasets, (3) specifying attribute structures, (4) conducting DCM analyses via Bayesian estimation, and (5) interpreting DCM results.

Step 1: Installing/loading R Packages

In this tutorial, the packages *xlsx*, *CDM*, and *R2jags* are required to load datasets, conduct DCM analyses, and implement Bayesian analyses, respectively. The R commands for installing (*install.packages*) and loading the packages are as follows:

```
install.packages(c ("xlsx", "CDM",
"R2jags"))

lapply(c("xlsx", "CDM", "R2jags"),
require, character.only = TRUE)
```

Step 2: Retrieving datasets

The R codes used to retrieve the datasets, Q-matrix and responses, in Excel are as follows:

```
dat_r <- read_excel("Data.xlsx")

dat_q <- read_excel("Q_matrix.xlsx")
```

Table 1. The Q-matrix

Item	Attribute 1	Attribute 2	Attribute 3
1	0	1	0
2	0	1	0
3	0	1	0
4	0	0	1
5	0	1	0
6	1	0	0
7	1	0	0
8	1	0	0
9	0	1	0
10	0	1	0
11	0	0	1
12	0	0	1

Step 3: Specifying hierarchical attribute structures

Four attribute structures with three attributes were examined in this demonstration; that is, one independent attribute structure served as the baseline model, and three were hierarchical attribute structures (i.e., linear, convergent, and divergent hierarchies). The independent attribute structure indicated no relationships among the three attributes. Figure 2 presents three hierarchical attribute structures among Attributes 1, 2 and 3, which correspond to linear, convergent, and divergent hierarchies, respectively.

When the attributes are in the hierarchical structure, possible attribute profiles are constrained, thereby reducing the number of possible attribute profiles and affecting the probability of classifying examinees into possible attribute profiles. The function *skillspace.hierarchy*, which is accompanied with two arguments (i.e., *B* and *skill.names*) in the *CDM* package, was used to define reduced attribute profiles for hierarchical structures among attributes. Argument *B* presents a string containing restrictions on the hierarchy, and *skill.names* defines attribute names. The following commands were used to generate the linear, convergent, and divergent attribute hierarchies, with the three attributes denoted as A1, A2, and A3, respectively (Figure 2).

```
linear <- skillspace.hierarchy (B =
"A1>A2>A3", skill.names =
paste0("A",1:3))
```

```
convergent <- skillspace.hierarchy (B
= "A1>A3 \n A2>A3", skill.names =
paste0("A",1:3))
```

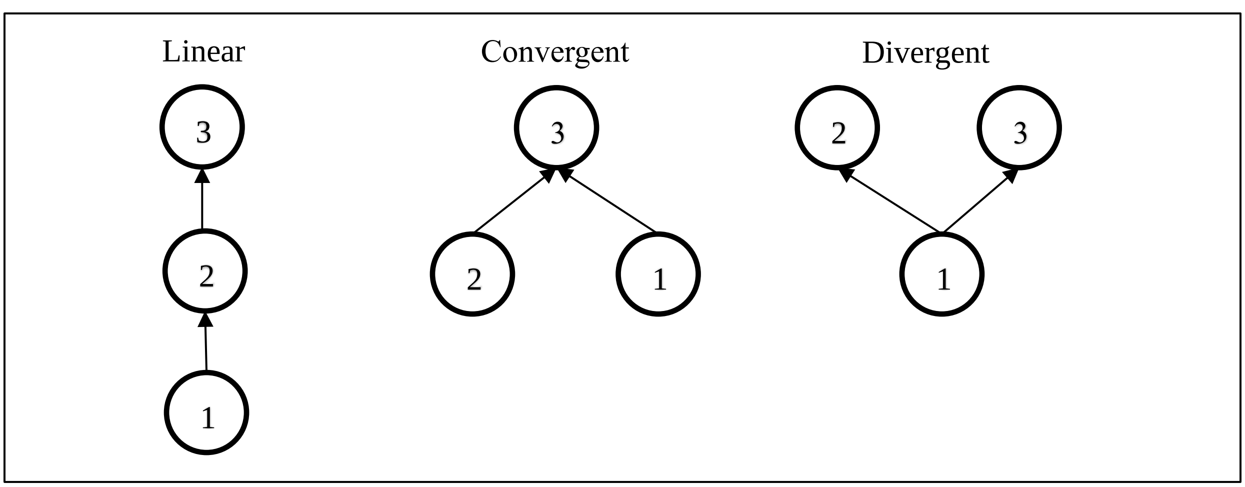
```
divergent<- skillspace.hierarchy (B =
"A1>A2 \n A1>A3", skill.names =
paste0("A",1:3))
```

Step 4: Conducting DCM analyses via Bayesian estimation

Before using *R2jas* to implement a Bayesian analysis for a DCM, we need to specify the DCM to be used. As mentioned earlier, we adopted the LCDM for the demonstration. Other DCMs (e.g., the DINA and DINO models) can also be applied with our tutorial to conduct a DCM analysis involving a hierarchical structure. Employing Bayesian estimation requires iterative simulations based on data to acquire parameter estimates.

To conduct a DCM analysis using the LCDM, we created a function called *LCDM_AH_Bayes*, which contains all calculations and estimations, as well as two components: the first component is related to the LCDM estimation and the settings of prior information for item parameters, and the second component is associated with the Bayesian analysis. The following commands were used to present the first component of *LCDM_AH_Bayes* with the three attributes (i.e., whole numbers and integers; fractions, decimals, and percentages; and data analysis and probability)

Figure 2. The investigated hierarchical attribute structures



```
1.  LCDM_AH_Bayes <- function(){
2.  for(i in 1:N){
3.          for(j in 1:J){
4.                  for (k in 1:K){w[i, j, k] <- alpha[i, k] * Q[j, k]}
5.                  lamda1_1[i, j] <- lamda1[j]*w[i, j, 1] + lamda2[j]*w[i, j, 2] + lamda3[j]*w[i,j, 3]
6.          logit(p[i, j]) <- lamda0[j] + lamda1_1[i, j]
7.          Y[i, j] ~ dbern(p[i, j])}
8.          for(k in 1:K) {alpha[i, j] <- profile[c[i], j]}
9.          c[n] ~ dcat(pai[1:C])}
10.         pai[1:C] ~ ddirch(delta[1:C])
11. for(j in 1:J) {
12.         lamda0[j] ~ dnorm(-1.096, 0.25)
13.         xlamda1[j] ~ dnorm(0, 0.25)%_% T(0,)
14.         xlamda2[j] ~ dnorm(0, 0.25)%_%T(0,)
15.         xlamda3[j] ~ dnorm(0, 0.25)%_%T(0,)
16.         lamda1[j] <- xlamda1[j]*Q[j, 1]
17.         lamda2[j] <- xlamda2[j]*Q[j, 2]
18.         lamda3[j] <- xlamda3[j]*Q[j, 3]}
```

Here, examinees, items, and attributes are denoted by i (i = 1, 2, …, N), j (j = 1, 2, …, J), and k (k = 1, 2, …, K), respectively. On the basis of Equation (1) of the LCDM, w[i, j, k] on Line 4 generates all parameters and indicators (i.e., $\mathbf{h}(\boldsymbol{\alpha}_i, \mathbf{q}_i)$) except for the intercept and lamda*k*[j] on Line 5, which represents the parameter for the main effect of attribute k on item j. The Q-matrix used in this study was a simple structure, indicating that each item measures a single attribute (Table 1), the number of the main effect parameters for each item was 1 (i.e., $\sum_{k=1}^{K} \mathrm{q}_{jk} = 1$), and no interaction effect parameters were used. The command on Line 6 is intended to create the probability of a response to item j of examinee i, as shown in Equation (1), which includes the intercept (i.e., lamda0[j]) and the main effect (i.e., lamda1_1[i, j]) parameters.

The command on Line 7 is the procedure for generating item responses. The commands on Lines 8 to 10 are the processes for generating attribute profiles. The commands on Lines 11 to 18 specify prior information for estimated parameters. Using the "%_%" prior to T(0,) on Lines 13 to 15 is meant to deal with the incompatibility issue that arises when Bayesian analysis is performed via the software *JAGS* (Plummer, 2017) in R.

For the second component of *LCDM_AH_Bayes*, Bayesian analysis was conducted, and the posterior predictive probability (ppp) value was calculated to evaluate the Bayesian estimation. A ppp close to .5 indicates that there are no systematic differences between a model and data, whereas a value close to 0 or 1 reflects inadequate model–data fit (Gelman et al., 2014). When researchers use other DCMs, there is no need to revise this part.

```
for (i in 1:N){
        for (j in 1:J){
                stat[i,j] <- pow(Y[i, j] - p[i, j], 2)/(p[i, j]*(1 - p[i, j]))
                Y_rep[i,j] ~ dbern(p[i,j])
                stat_rep[i,j]<- pow(Y_rep[i, j] - p[i, j],2)/(p[i, j]*(1 - p[i, j]))}}
        sum_stat <- sum(stat[1:N, 1:J])
        sum_statrep <- sum(stat_rep[1:N, 1:J])
        ppp <- step(sum_statrep-sum_stat)}
```

Before a Bayesian analysis is conducted, information on the data and estimated parameters of interest should be specified. The following commands were used for this purpose:

```
N <- nrow(dat_r)
J <- nrow(dat_q)
K <- ncol(dat_q)
C <- nrow(linear)
```

```
5.  delta <- rep(1, C)
6.  jags.data <- list("N", "J", "K",
    "Y", "Q", "C", "profile", "delta")
```

The command on Lines 1 to 4 represents the sample size (N), number of items (J), number of attributes (K), and number of attribute profiles (C), respectively, and the command on Line 5 is used for the first component of *LCDM_AH_Bayes*. Finally, the command on Line 6 represents data information. In Bayesian analysis, a preliminary study is required to examine the convergence of parameter estimations. The command for a convergence check is listed below. The parameters we were interested in included the intercept (lamda0), three parameters for the main effects of the three attributes (lamda1, lamda2, and lamda3), and the proportion of attribute profiles (pai).

```
pre.jags.parameters <-
c("lamda0","lamda1", "lamda2",
"lamda3", "pai")
```

The commands below were used to perform the Bayesian analysis using *R2jags*.

```
1.  jags.inits <- NULL
2.  pre.sim_Model1 <-
    jags(data=jags.data,
    inits=jags.inits,
    parameters.to.save=pre.jags.parame
    ters, model.file=LCDM_AH_Bayes,
    n.chains=2, n.iter=2000,
    n.burnin=1000,n.thin=1,DIC=TRUE)
```

The command on Line 1 indicates that we did not set initial values for parameter estimation. Establishing appropriate initial values can save time on convergence during parameter estimation. The command on Line 2 was executed to activate the Bayesian analytical procedure. In the function of *jags*, we used two chains (n.chains = 2), 2,000 iterations per chain (n.iter = 2000), and the first 1,000 as burn-in (n.burnin = 1000). This means that after 2,000 iterations were completed, the first 1,000 iterations in each chain were excluded so that later 1,000 iterations could be used to calculate the estimates of the parameters.

After 2,000 iterations, we examined convergence using $\hat{R}$ (Gelman & Rubin, 1992), executed as the following command:

```
1.  R_convergence <-
    sum(pre.sim_Model1$BUGSoutput$summ
    ary[ , 8]>= 1.2)== 0
2.  if(R_convergence==1){pre.sim_Model
    1$n.iter}
3.  if(R_convergence==0){pre.sim_Model
    1.c <- autojags(sim_Model1,
    Rhat=1.2,n.update=30)
    pre.sim_Model1.c$n.iter}
```

If all $\hat{R}$ values were fulfilled in accordance with the requirement of $\hat{R} \leq 1.2$, then all parameter estimations converged (Line 2); otherwise, the estimated parameters were automatically updated until all $\hat{R}$ values were less than 1.2 (Line 3). The results of the four attribute hierarchy models showed that this criterion was satisfied for all the parameter estimates, indicating that all such estimations converged. In this study, 2,000 iterations were needed to address the criterion of $\hat{R} \leq 1.2$ for all the estimated parameters across all the attribute hierarchy models.

The following commands were used to perform a Bayesian analysis with the linear attribute hierarchy. Similar settings were implemented, but we changed the number of iterations to 10,000 and the burn-in to 5,000 for a more stable estimation.

```
1.  fin.jags.parameters <-
    c("lamda0","lamda1", "lamda2",
    "lamda3", "pai","c","ppp")
2.  fin.sim_Model1 <-
    jags(data=jags.data,
    inits=jags.inits,
    parameters.to.save=fin.jags.parame
    ters, model.file=LCDM_AH_Bayes,
    n.chains=2,n.iter=10000,
    n.burnin=5000,n.thin=1, DIC=TRUE)
```

Step 5: Interpreting DCM analysis results

The following command was used to obtain the Bayesian analysis results of the linear hierarchy model, including the deviance information criterion (DIC), the ppp, the proportion of attribute profiles, and the item parameter of the model.

```
sim_Model1.result <-
sim_Model1$BUGSoutput
```

To choose the best-fit attribute hierarchy for the 2007 TIMSS mathematics achievement data, the fit indices of the DIC and ppp were used for model comparisons. Table 2 shows the model fit with respect to the relative (e.g., DIC) and absolute (e.g., ppp) fit indices and the number of attribute profiles across the four hierarchies. Note that the numbers of item parameters were identical (i.e., 12 intercepts and 12 main effect parameters) across the four hierarchies because the same DCM (i.e., the LCDM) was conducted. However, the numbers of attribute profiles differed across the models because they were characterized by varying structures of attribute hierarchy. As shown in Table 2, the linear attribute hierarchy had the smallest DIC (7742) and the minimum number of attribute profiles (four attribute profiles, i.e., 000, 100, 110, and 111). The absolute model fit of ppp for the linear, convergent, and for divergent hierarchies was approximately .38, which was slightly higher than that for the independent model (.37). These ppp values were not close enough to the expected value of .5.

Table 3 presents the classification proportion of examinees for each attribute profile. Most of the examinees were consistently classified into the (000) and (111) attribute profiles across the four attribute hierarchies, indicating that approximately 85% or more of the students were likely to have mastered no attributes (i.e., 000) or all attributes (i.e., 111), regardless of attribute hierarchy. The standard errors of estimation were sufficiently close (i.e., approximately .03) across the four attribute hierarchies. The proportions of attribute profiles with standard errors did not provide statistically significant evidence for choosing the best-fitting hierarchy in this study.

Table 4 displays the estimates of item parameters and corresponding standard errors of the linear attribute hierarchy. The tables that illustrate these parameter estimates for the other three attribute hierarchies can be found in the Appendix. On average, the linear attribute hierarchy yielded slightly smaller standard errors of item parameter estimates than those generated by the other three attribute hierarchies. These findings, including the smallest DIC, similar ppp values, smallest number of attribute profiles, and smallest standard errors, consistently indicate that the linear attribute hierarchy would be the best-fitting model to describe the attribute hierarchy structure of the 2007 TIMSS mathematics test.

Discussion

DCMs have been extensively employed in education assessments because cognitive profiles that are generated via DCM analyses can provide teachers or stakeholders with more diagnostic feedback with respect to a set of fine-grained latent student attributes, thereby enabling them to develop the tailored or remedial training or curricula that are necessary to improve teaching and learning. An important

Table 2. Model fit indices across various models of attribute hierarchy

Hierarchy	DIC	ppp	Number of Attribute Profiles
Independent	8749	0.374	8
Linear	7742	0.378	4
Convergent	8131	0.385	5
Divergent	7963	0.381	5

Note: DIC = deviance information criterion, ppp = posterior predictive probability

Table 3. Proportions of attribute profiles across various models of attribute hierarchy

Attribute Profile	Independent	Linear	Convergent	Divergent
(0,0,0)	0.457 (0.044)	0.513 (0.040)	0.495 (0.041)	0.509 (0.040)
(1,0,0)	0.022 (0.017)	0.030 (0.024)	0.027 (0.020)	0.028 (0.021)
(0,1,0)	0.022 (0.017)	-	0.026 (0.019)	-
(1,1,0)	0.027 (0.020)	0.036 (0.025)	0.038 (0.027)	0.039 (0.026)
(0,0,1)	0.032 (0.023)	-	-	-
(1,0,1)	0.011 (0.010)	-	-	0.014 (0.012)
(0,1,1)	0.033 (0.023)	-	-	-
(1,1,1)	0.397 (0.040)	0.421 (0.039)	0.414 (0.041)	0.411 (0.041)

Note: The values in parentheses are standard errors of estimation. The dash (-) indicates that an attribute profile is inapplicable to a given model.

Table 4. Item parameter estimates and standard errors of the linear hierarchy model

Item	λ_0	$\lambda_{1,(1)}$	$\lambda_{1,(2)}$	$\lambda_{1,(3)}$
1	1.039 (0.143)	-	1.260 (0.298)	-
2	0.547 (0.133)	-	1.724 (0.312)	-
3	0.735 (0.131)	-	0.592 (0.236)	-
4	–1.103 (0.178)	-	-	3.221 (0.496)
5	–1.272 (0.183)	-	2.805 (0.292)	-
6	0.728 (0.141)	1.791 (0.34)	-	-
7	0.710 (0.148)	1.822 (0.325)	-	-
8	–1.656 (0.254)	3.020 (0.33)	-	-
9	–0.310 (0.128)	-	0.796 (0.202)	-
10	–0.558 (0.144)	-	1.451 (0.221)	-
11	1.237 (0.149)	-	-	1.462 (0.495)
12	0.751 (0.139)	-	-	1.944 (0.488)

Note: λ_0 = intercept parameter; $\lambda_{1,(k)}$ = main effect parameter associated with attribute *k*. The values in parentheses are standard errors of estimation. The dash (-) indicates that an attribute profile is inapplicable to a given model.

consideration, however, is that misspecified attribute structures may result in harmful effects on parameter estimation and classification accuracy (e.g., Liu, 2018; Liu et al., 2017; Templin & Bradshaw, 2014; Templin et al., 2008), ultimately causing inappropriate decision making. Hence, with a rise in the analysis of educational or psychological data via DCMs with various attribute structures, a walk-through demonstration can help applied researchers easily understand how such an analysis is carried out. The present study performed Bayesian estimation because it offers better differentiation when conducting DCM analyses with different attribute structures and increases parameterization flexibility (e.g., addressing the issue of small sample, integrating a trustworthy prior to enhance parameter estimation, and eliminating the need for specialized analytical software for Bayesian estimation). Furthermore, the employed R software is freely available, affording everyone convenient use and practice with their own data. We demonstrated how R can be used to implement a DCM analysis with various hierarchical attribute structures and how to select an appropriate attribute structure to describe data using the DIC, ppp, proportion of attribute profiles, and estimated parameters.

This tutorial is aimed at increasing researchers' studying and/or applying in the fields of CDM and/or Bayesian estimation by using R, especially for those who are interested and unfamiliar with these three fields. Thus, we recommend that researchers and practitioners can use our codes directly, and they need only to specify their attribute structure and LCDM in the code. Regarding the LCDM, they should pay attention to whether the Q-matrix is a simple structure. If the Q-matrix is a simple structure like our demonstration, there is no need to specify the interaction terms between attributes. However, if one item measures two or more attributes, they should include all possible interaction terms in LCDM.

References

Birenbaum, M., Tatsuoka, C., & Yamada, T. (2004). Diagnostic assessment in TIMSS-R: Between-countries and within-country comparisons of eighth graders' mathematics performance. *Studies in Educational Evaluation*, *30*, 151–173. https://doi.org/10.1016/j.stueduc.2004.06.004

Chen, Y.-H. (2012). Cognitive diagnosis of mathematics performance between rural and urban students in Taiwan. *Assessment in Education*, *19*(2), 193-209. https://doi.org/10.1080/0969594X.2011.560562

Darling-Hammond, L., Barron, B., Pearson, P. D., Schoenfeld, A. H., Stage, E. K., Zimmerman, T. D., & Tilson, J. L. (2015). *Powerful learning: What we know about teaching for understanding*. Wiley.

de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, *76*, 179-199. https://doi.org/10.1007/s11336-011-9207-7

DiBello, L. V., Stout, W., & Roussos, L. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In P. Nichols, S. Chipman, & R. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 361-390). Lawrence Erlbaum.

Dogan, E., & Tatsuoka, K. K. (2008). An international comparison using a diagnostic testing model: Turkish students' profile of mathematical skills on TIMSS-R. *Educational Studies in Mathematics*, *68*, 263–272. https://doi.org/10.1007/s10649-007-9099-8

Entwistle, N., & Ramsden, P. (2015). *Understanding student learning (Routledge revivals)*. Routledge.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, *7*, 457-511. https://doi.org/10.1214/ss/1177011136

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis*, *3rd edition*. CRC Press.

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

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. https://doi.org/10.1007/s11336-008-9089-5

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. https://doi.org/10.1177/01466210122032064

Lee, Y.-S., Johnson, M. S., Park, J. Y., Sachdeva, R., Zhang, J., and Waldman, M. (2013 April). *A MDS Approach for investigating student's cognitive weakness and strength on the TIMSS 2007 mathematics assessment.* AERA annual meeting. San Francisco, CA.

Lee, Y.-S., Park, Y. S., & Taylan, D. (2011). A cognitive diagnostic modeling of attribute mastery in Massachusetts, Minnesota, and the U.S. national sample using the TIMSS 2007. *International Journal of Testing*, *11*, 144-177. https://doi.org/10.1080/15305058.2010.534571

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*, 205-237. https://doi.org/10.1111/j.1745-3984.2004.tb01163.x

Liu, R. (2018). Misspecification of attribute structure in diagnostic measurement. *Educational and Psychological Measurement*, *78*, 605-634. https://doi.org/10.1177/0013164417702458

Liu, R., Huggins-Manley, A. C., & Bradshaw, L. (2017). The impact of Q-matrix designs on diagnostic classification accuracy in the presence of attribute hierarchies. *Educational and Psychological Measurement*, *77*, 220-240. https://doi.org/10.1177/0013164416645636

Ma, W. (2019). A diagnostic tree model for polytomous responses with multiple strategies. *British Journal of Mathematical and Statistical Psychology, 72*, 61-82. https://doi.org/10.1111/bmsp.12137

Plummer, M. (2003). JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. *The 3rd International Workshop on Distributed Statistical Computing*, *124*, 1–8. Retrieved from http://www.ci.tuwien.ac.at/Conferences/DSC-2003/

Plummer, M. (2017). *JAGS Version 4.3.0 user manual.* Retrieved from: https://people.stat.sc.edu/hansont/stat740/jags_user_manual.pdf

R Core Team. (2012). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing Vienna Austria. Retrieved from http://www.R-project.org/

Robitzsch, A., Kiefer, T., George, A. C., & Uenlue, A. (2019). *CDM: Cognitive diagnosis modeling. R package version 7.3-17*. Retrieved from https://CRAN.R-project.org/package=CDM

Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. Guilford Press.

Su, Y.-S., & Yajima, M. (2020). *R2jags: Using R to run 'JAGS'*. R package version 0.6-1. Retrieved from https://CRAN.R-project.org/package=R2jags

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, *20*, 345-354. https://doi.org/10.1111/j.1745-3984.1983.tb00212.x

Tatsuoka, K. K., Corter, J., & Tatsuoka, C. (2004). Patterns of diagnosed mathematical content and process skills in TIMSS-R across a sample of 20 countries. *American Educational Research Journal*, *41*, 901–926. https://doi.org/10.3102/00028312041004901

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, *11*(3), 287-305. https://doi.org/10.1037/1082-989X.11.3.287

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*, 559-574. https://doi.org/10.1177/0146621607300286

Templin, J., & Bradshaw, L. (2014). Hierarchical diagnostic classification models: A family of models for estimating and testing attribute hierarchies. *Psychometrika*, *79*, 317-339. https://doi.org/10.1007/s11336-013-9362-0

von Davier, M. (2005). *A general diagnostic model applied to language testing data*. ETS Research Report. No. RR-05-16. Educational Testing Service.

Zhang, X., & Wang, J. (2020). On the sequential hierarchical cognitive diagnostic model. *Frontiers in Psychology,* *11*:579018. https://doi:10.3389/fpsyg.2020.579018

Citation:

Hsu, C.-L., Chen, Y.-H., &Wu, Y.-J. (2023). Using a Bayesian estimation to examine attribute hierarchies of the 2007 TIMSS mathematics test : A demonstration using R packages. *Practical Assessment, Research, & Evaluation*, 28(11). Available online: https://scholarworks.umass.edu/pare/vol28/iss1/11/

Corresponding Author:

Chia-Ling Hsu
Hong Kong Examinations and Assessment Authority
Email: clhsu [at] hkeaa.edu.hk

Appendix

Table A1. Item parameter estimates and standard errors for the independent attribute model

Item	λ_0	$\lambda_{1,(1)}$	$\lambda_{1,(2)}$	$\lambda_{1,(3)}$
1	1.017 (0.147)		1.256 (0.293)	
2	0.503 (0.139)		1.752 (0.311)	
3	0.724 (0.138)		0.588 (0.246)	
4	-1.352 (0.256)			3.282 (0.508)
5	-1.443 (0.250)		2.998 (0.342)	
6	0.773 (0.142)	1.820 (0.358)		
7	0.743 (0.143)	1.903 (0.364)		
8	-1.633 (0.246)	3.258 (0.421)		
9	-0.318 (0.136)		0.774 (0.217)	
10	-0.595 (0.152)		1.462 (0.228)	
11	1.149 (0.162)			1.578 (0.471)
12	0.638 (0.158)			2.06 (0.455)

Note: λ_0 = intercept parameter; $\lambda_{1,(k)}$ = main effect parameter associated with attribute *k*. The values in parentheses are standard errors of estimation. The dash (-) indicates that an attribute profile is inapplicable to a given model.

Table A2. Item parameter estimates and standard errors for the convergent attribute hierarchy model

Item	λ_0	$\lambda_{1,(1)}$	$\lambda_{1,(2)}$	$\lambda_{1,(3)}$
1	1.016 (0.147)		1.258 (0.295)	
2	0.512 (0.134)		1.728 (0.309)	
3	0.729 (0.136)		0.579 (0.236)	
4	-1.079 (0.182)			3.226 (0.509)
5	-1.387 (0.232)		2.893 (0.323)	
6	0.740 (0.141)	1.798 (0.338)		
7	0.723 (0.144)	1.816 (0.335)		
8	-1.641 (0.249)	3.056 (0.344)		
9	-0.328 (0.132)		0.797 (0.207)	
10	-0.597 (0.15)		1.468 (0.225)	
11	1.230 (0.152)			1.572 (0.538)
12	0.748 (0.134)			2.057 (0.556)

Note: λ_0 = intercept parameter; $\lambda_{1,(k)}$ = main effect parameter associated with attribute *k*. The values in parentheses are standard errors of estimation. The dash (-) indicates that an attribute profile is inapplicable to a given model.

Table A3. Item parameter estimates and standard errors for the divergent attribute hierarchy model

Item	λ_0	$\lambda_{1,(1)}$	$\lambda_{1,(2)}$	$\lambda_{1,(3)}$
1	1.047 (0.145)		1.265 (0.303)	
2	0.552 (0.132)		1.743 (0.326)	
3	0.739 (0.134)		0.599 (0.246)	
4	-1.107 (0.184)			3.220 (0.530)
5	-1.270 (0.190)		2.871 (0.303)	
6	0.722 (0.145)	1.796 (0.334)		
7	0.709 (0.143)	1.799 (0.321)		
8	-1.653 (0.240)	3.002 (0.333)		
9	-0.306 (0.128)		0.800 (0.210)	
10	-0.546 (0.143)		1.453 (0.223)	
11	1.227 (0.150)			1.537 (0.515)
12	0.736 (0.140)			2.031 (0.559)

Note: λ_0 = intercept parameter; $\lambda_{1,(k)}$ = main effect parameter associated with attribute *k*. The values in parentheses are standard errors of estimation. The dash (-) indicates that an attribute profile is inapplicable to a given model.