Proceedings of the 89th Annual International Meeting of the Psychometric Society, Prague, Czech Republic (2024).

The Impact of Attribute Hierarchies' Distribution on Diagnostic Classification Accuracy

Chia-Ling Hsu*† (ORCID: https://orcid.org/0000-0002-4267-0980) and Yi-Jhen Wu‡

† Hong Kong Examinations and Assessment Authority, Hong Kong SAR, China

‡ Center for Research on Education and School Development, Technical University Dortmund, Germany

*Corresponding author. Email: clhsu@hkeaa.edu.hk

Abstract

Cognitive diagnostic models (CDMs) give detailed information about how well examinees' grasp a set of fine-grained, discrete, latent skills/attributes. This information allows researchers and teachers to tailor instruction and craft cost-effective interventions to improve student learning. While learning, a student typically masters lower-level skills before a higher-level skill, which suggests four hierarchical attribute structures: linear, convergent, divergent, and unstructured (Leighton et al., 2024). Past studies assumed that all students within a sample have the same hierarchical structure (e.g., linear structure for primary school students' learning of arithmetic). However, students' learning processes can vary widely and yield different hierarchical attribute structures. Recognizing this possibility, this study ran simulations to test how well cognitive diagnostic models classified students across hierarchical attribute structures. The findings revealed that the distribution of these structures impacted classification accuracy. More candidate hierarchical attribute structures for classifying students yielded greater accuracy.

Keywords: cognitive diagnostic model, attribute structure, classification accuracy, misspecification, diagnostic measurement

1. Introduction

Cognitive diagnostic models (CDMs; Rupp et al., 2010) analyze responses to test questions/items to reveal whether each person has mastered specific fine-grained latent skills (also known as attributes). This information helps teachers adapt their curricula, course materials, and teaching methods to meet their students' needs. Initially, CDMs sought to identify which skills a student had mastered, but now, they also aim to grasp how students learn these skills in sequence, which has driven the development of hierarchical CDMs (HCDMs) to map out these learning paths (Leighton et al., 2004).

Prior studies examined the effects of misspecified hierarchical structures on parameter estimation and classification accuracy (e.g., Liu, 2018; Liu et al., 2016; Templin et al., 2008). Although past models often assumed that all students followed the same learning path, their actual paths could differ (mixture of hierarchical attribute structures), which led to inaccurate classifications. To address this shortcoming, we ran simulations to assess how mixed attribute hierarchies within a sample affect classification accuracy in CDMs. After a brief overview of CDMs and the hierarchical structure of attributes, we lay out the simulation design. Then, we report our findings, and discuss their implications.

2. Cognitive diagnostic models (CDMs)

In CDMs, a Q-matrix defines the links between items and attributes through an item-to-attribute mapping $J \times K$ matrix (Tatsuoka, 1983). J is the number of items (j=1,…, J), and K is the number of attributes (k=1,…, K). Each Q-matrix element q_{jk} specifies whether the k^{th} attribute is required to solve the j^{th} item correctly: if yes, q_{jk} = 1, otherwise, q_{jk} = 0. Also, the vector $\boldsymbol{\alpha}_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{iK})$ captures individual mastery of each attribute (also known as *attribute profile*). If the i^{th} individual has mastered the k^{th} attribute, α_{ik} = 1; otherwise, α_{ik} = 0.

In this study, we used the *deterministic inputs, noisy "AND" gate* (DINA) model (Junker & Sijtsma, 2001) to examine how mixed attribute hierarchies affect classification accuracy. The DINA model assumes that correctly answering an item requires mastery of all its specified attributes. In this framework, a latent response vector of the j^{th} item for the i^{th} individual is defined as $\boldsymbol{\eta}_i = (\eta_{i1}, \eta_{i2}, \dots, \eta_{ij})$, where $\eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}}$. If the i^{th} individual has mastered all required attributes for the j^{th} item, η_{ij} = 1; otherwise, η_{ij} = 0.

The DINA model has two key parameters: slipping (s_j) and guessing (g_j). Slipping is the probability of an incorrect response by individuals who have mastered all the required attributes for the j^{th} item. Conversely, guessing is the probability of a correct response by individuals who have not mastered all of these attributes for the j^{th} item. For the attribute vector $\boldsymbol{\alpha}_i$, the probability of individual i correctly answer item j (X_{ij}) is:

$$P(X_{ij} = 1 | \boldsymbol{\alpha}_i) = g_j^{1-\eta_{ij}} (1 - S_j^{\eta_{ij}}) \tag{1}$$

We chose the DINA model for its simplicity and ease of interpretation (de la Torre, 2009), though other CDMs can also be used to examine mixed hierarchical structures of attributes.

3. Hierarchical structure of attributes

Attributes that students learn are often hierarchically correlated (e.g., Crowley, 1987; Mayer, 1996). For instance, in arithmetic, students must grasp basic concepts in addition and subtraction as prerequisites for learning advanced operations like multiplication and division. Leighton et al. (2004) proposed four types of hierarchical attribute structures: *linear*, *convergent*, *divergent*, and *unstructured* (see Figure 1). In the linear structure, failure to master the first attribute (A1) inhibits mastery of the subsequent attributes (A2–A5). The convergent structure offers a choice: master either one of the attributes (A3 or A4) or both to achieve mastery of A5. In the divergent structure, a person must master all three preceding attributes (A1, A2, and A3) before mastering either one of the subsequent attributes (A4, or A5). Lastly, for the unstructured case, mastery of one attribute (A1) enables mastery of any of subsequent attributes (A2, A3, A4, or A5).

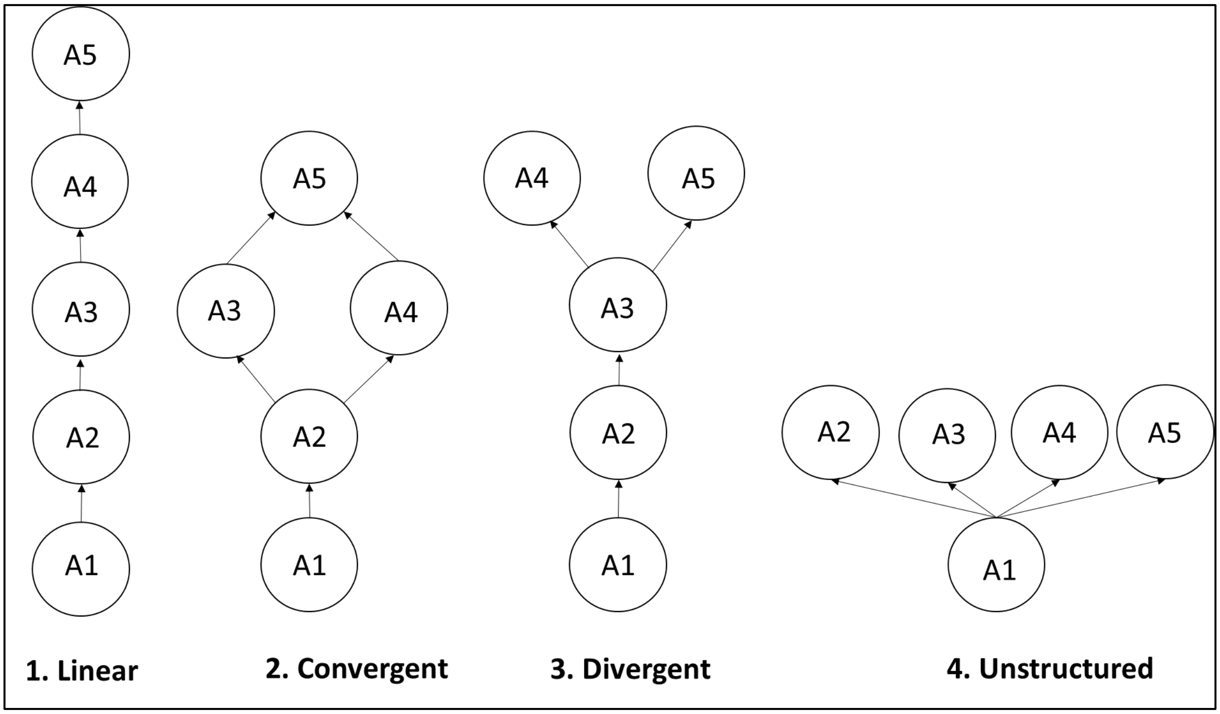

Figure 1. Four Hierarchical Structures Using Five Attributes

Note. A means attribute.

Allowing different hierarchical structures limits the number of attribute profiles available to recover individuals' mastery statuses (rather than all possible attribute profiles, see Table 1). For example, the linear structure is the most restrictive, yielding only six possible attribute profiles. By contrast, the unstructured one is the least restrictive and yields 17 attribute profiles. These hierarchical structures share possible attribute profiles to varying degrees. As shown in Table 1, all attribute profiles (i.e., six attribute profiles) in the linear structure are also present in the convergent, divergent, and unstructured structures. However, nine attribute profiles in the unstructured structure do not appear in the linear, convergent, or divergent structures. This suggests that misclassifying attribute order (e.g., assigning lower-level attributes to a higher-level structure) or structures (e.g., fitting a linear structure to unstructured data) reduces classification accuracy (e.g., Liu, 2018; Liu et al., 2016; Templin et al., 2008).

Table1. The Attribute Profiles Across Various Hierarchical Structures Using Five Attributes

1. Linear (*P*=6)	2. Convergent (P=7)	3. Divergent (*P*=7)	4. Unstructured (*P*=17)
(0,0,0,0,0)	(0,0,0,0,0)	(0,0,0,0,0)	(0,0,0,0,0)

(1,0,0,0,0)	(1,0,0,0,0)	(1,0,0,0,0)	(1,0,0,0,0)
(1,1,0,0,0)	(1,1,0,0,0)	(1,1,0,0,0)	(1,1,0,0,0)
(1,1,1,0,0)	(1,1,1,0,0)	(1,1,1,0,0)	(1,1,1,0,0)
(1,1,1,1,0)	(1,1,1,1,0)	(1,1,1,1,0)	(1,1,1,1,0)
(1,1,1,1,1)	(1,1,1,1,1)	(1,1,1,1,1)	(1,1,1,1,1)
	(1,1,0,1,0)		(1,1,0,1,0)
		(1,1,1,0,1)	(1,1,1,0,1)
			(1,0,1,0,0)
			(1,0,0,1,0)
			(1,0,1,1,0)
			(1,0,0,0,1)
			(1,1,0,0,1)
			(1,0,1,0,1)
			(1,0,0,1,1)
			(1,1,0,1,1)
			(1,0,1,1,1)

Note. *P* means the number of attribute profiles, and grey colors mean that an attribute profile presented in other hierarchical structures.

4. Simulation study

Our simulation study explores the impacts of heterogeneous hierarchical structures on classification accuracy. Specifically, we introduced three mixed proportions of hierarchical structure combinations (0.5/0.5, 0.2/0.8, and 0.8/0.2) among all possible pairs of six heterogeneous hierarchical structures within the four hierarchies in Figure 1 (e.g., linear vs. convergent, linear vs. divergent, and linear vs. unstructured). This yielded 18 types of hierarchical structure mixtures (see Table 2). In the generation process, each hierarchical structure was represented by three categories within the mixtures: equivalent proportions, minor proportions, and major proportions. For example, Hierarchical Structure Mixture 1 featured an equivalent proportion of linear and convergent hierarchies, whereas Hierarchical Structure Mixture 2 had a minor proportion (i.e., 0.2) from the linear hierarchy and a major proportion (i.e., 0.8) from the divergent hierarchy.

Table 2. The Hierarchical Structure Mixtures

Hierarchical Structure Mixture	Mixed Proportion	Hierarchy 1	Hierarchy 2
1	0.5/0.5	Linear	Convergent
2	0.2/0.8	Linear	Convergent
3	0.8/0.2	Linear	Convergent
4	0.5/0.5	Linear	Divergent
5	0.2/0.8	Linear	Divergent
6	0.8/0.2	Linear	Divergent
7	0.5/0.5	Linear	Unsturctured
8	0.2/0.8	Linear	Unsturctured
9	0.8/0.2	Linear	Unsturctured
10	0.5/0.5	Convergent	Divergent

11	0.2/0.8	Convergent	Divergent
12	0.8/0.2	Convergent	Divergent
13	0.5/0.5	Convergent	Unsturctured
14	0.2/0.8	Convergent	Unsturctured
15	0.8/0.2	Convergent	Unsturctured
16	0.5/0.5	Divergent	Unsturctured
17	0.2/0.8	Divergent	Unsturctured
18	0.8/0.2	Divergent	Unsturctured

We investigated the effects of heterogeneous hierarchical distributions on classification accuracy. To minimize the influence of the selected CDM, Q-matrix design, and item quality, we used a basic CDM (i.e., the DINA model), two simple Q-matrix structures, and high-quality items. Accordingly, we fixed both the *g*- and *s*-parameters to 0.1 in the DINA model. In the two Q-matrix structures, one had 10 items measuring five attributes and the other had 20 items measuring five attributes. Each item measured a single attribute. We generated 300 examinees. Each examinee had a 50% chance of mastering each attribute independently. This specification yielded 36 conditions: 2 (Q-matrix structure) × 3 (mixed proportion) × 6 (hierarchical structure combination) = 36. Each condition was replicated 30 times.

We fitted each of the 18 mixtures to the four hierarchies in Figure 1 separately, running the simulations with Bayesian estimation via the R package. We assessed classification accuracy by the proportion of examinees with correctly classified attribute profiles:

$$\frac{\sum_{i=1}^{N} I(\hat{\boldsymbol{\alpha}}_i \boldsymbol{\alpha}_i)}{N}. \quad (2)$$

If $\hat{\boldsymbol{\alpha}}_i = \boldsymbol{\alpha}_i$, the indicator function $I(\hat{\boldsymbol{\alpha}}_i \boldsymbol{\alpha}_i)$ equals 1; otherwise, it equals 0. *N* is the total number of examinees.

5. Simulation result

Figures 2 and 3 show the classification accuracies across 36 simulation conditions, fitting the linear, convergent, divergent, and unstructured hierarchies. As anticipated, the classification accuracy rates under the 20×5 Q-matrix were higher than those observed for the 10×5 Q-matrix. This pattern held consistent across different mixture proportions and hierarchical structure compositions. The superior performance of the 20×5 Q-matrix can be attributed to its greater informational capacity; the larger number of items facilitates a more precise estimation of examinees' attribute profiles compared to the smaller items in the 10×5 Q-matrix.

Regardless of the Q-matrix design, mixture proportion, or hierarchical structure composition, classification accuracy declined when ignoring hierarchical structures with greater heterogeneity (by estimation with a restricted hierarchy). This effect was most stark when a minimally restricted hierarchy constituted more than a minor proportion of the overall hierarchical structure. Sepcifically, when using predefined, restricted hierarchical structures (e.g., convergent), classification accuracy was lower with an unstructured hierarchy within the mixture (e.g., convergent vs. unstructured)—especially with equivalent or major proportions—than without it.

Conversly, the decline in classification accuracy diminished when estimating with a predefined, minimally restricted hierarchy (e.g., unstructured). That is, estimation with a predefined unstructured hierarchical structure, which offers many attribute profiles, mitigated this decline. For instance, within the 10×5 Q-matrix, the classification accuracy rates for three mixed proportions of linear and unstructured hierarchies (i.e., 0.2/0.8, 0.5/0.5, and 0.8/0.2) were: linear (.42, .58, .72), convergent (.47, .58, .73), divergent (.45, .60, .73), and unstructured (.70, .75, .77).

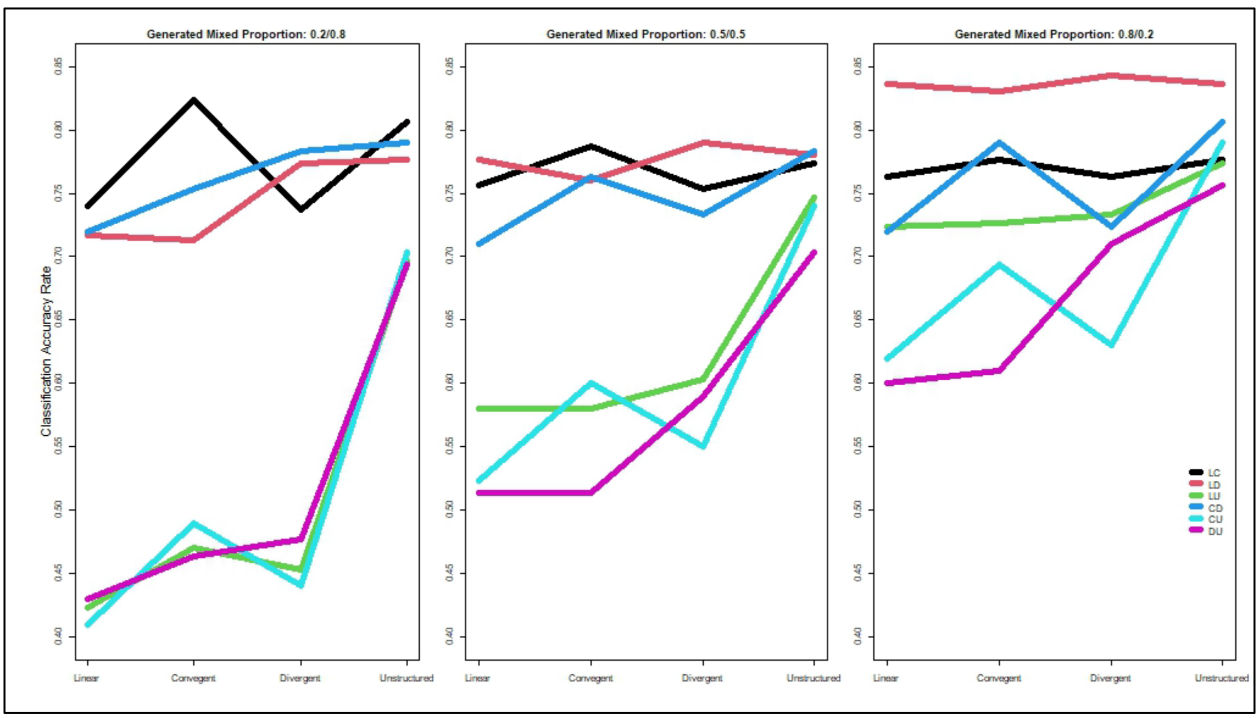

Figure 2. Classification Accuracy Under 10×5 Q-matrix

Note. LC = linear and convergent hierarchies, LD = linear and divergent hierarchies, LU = linear and unstructured hierarchies, CD = convergent and divergent hierarchies, CU = convergent and unstructured hierarchies, DU = divergent and unstructured hierarchies, and Generated Mixed Proportion = proportions generated from two hierarchical structures (e.g., Generated Mixed Proportion**:** 0.2/0.8 with LC, indicating that examinees' attribute structures were generated 20% from the linear hierarchy and 80% from the divergent hierarchy**.**

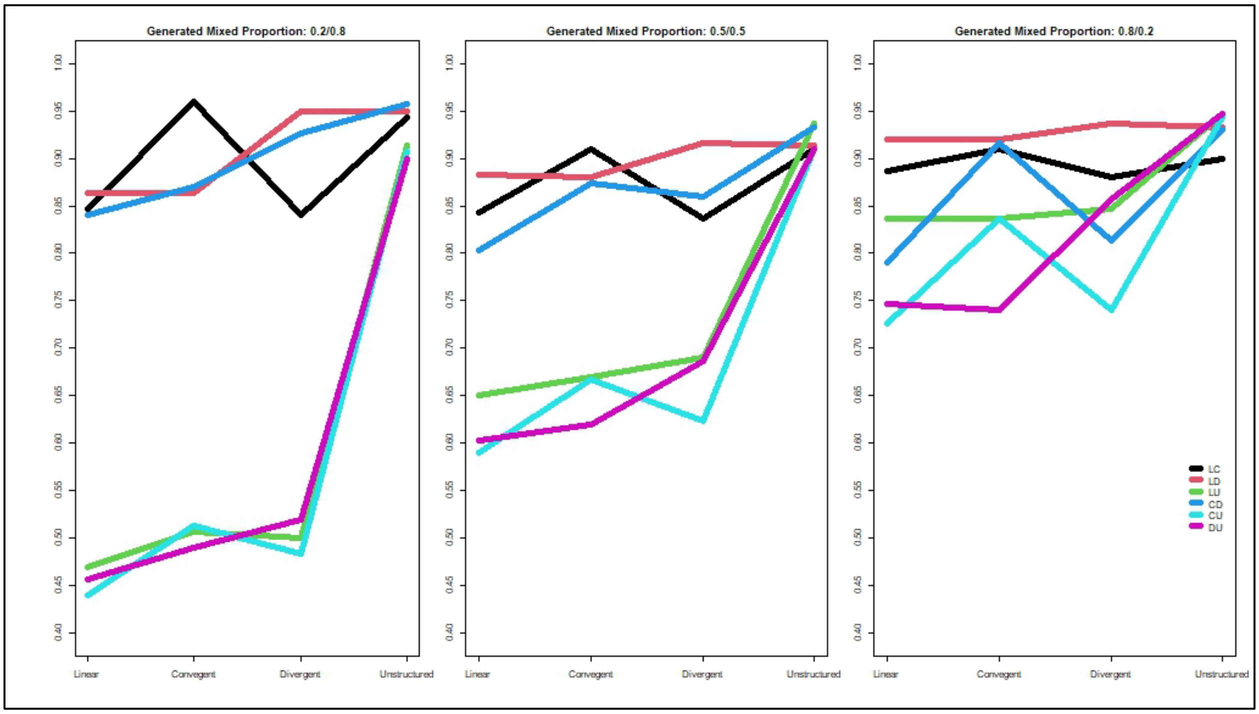

Figure 3. Classification Accuracy Under 20×5 Q-matrix

Note. LC = linear and convergent hierarchies, LD = linear and divergent hierarchies, LU = linear and unstructured hierarchies, CD = convergent and divergent hierarchies, CU = convergent and unstructured hierarchies, DU = divergent and unstructured hierarchies, and Generated Mixed Proportion = proportions generated from two hierarchical structures (e.g., Generated Mixed Proportion**:** 0.2/0.8 with LC, indicating that examinees' attribute structures were generated 20% from the linear hierarchy and 80% from the divergent hierarchy**.**

6. Conclusion

This study examined how ignoring heterogeneity within a hierarchical structure affected classification accuracy in CDM. Simulation studies of various degrees of heterogeneity within hierarchical structures showed that they affect classification accuracy. Neglect of greater heterogeneity reduced accuracy. This decline was especially severe when less restrictive hierarchies comprised substantial proportions of the structure and/or when unstructured hierarchies accompanied restricted ones. Conversely, using predefined, less restricted hierarchies (e.g., unstructured hieararchies) mitigated the decline. The underlying reason may stem from the limited number of attribute profiles available for estimating individuals' mastery statuses. Specifically, restricted hierarchies offer fewer attribute profiles, thereby constraining the recovery capacity in addressing greater heterogeneity. In contrast, less restricted hierarchies provide a wider range of attribute profiles, enhancing the ability to accommodate heterogeneity and improving recovery potential.

These findings underscore how addressing heterogeneity and aligning structure boosts classification accuracy. Moreover, heterogeneity is inherently complex and unpredictable. Advanced skills (e.g., higher-order cognitive abilities) are typically more intricate than basic ones (e.g., lower-order abilities). Consequently, adopting a minimally restricted hierarchy, such as an unstructured hierarchy, might be optimal. One practical suggestion is to first identify one or two fundamental prerequisite skills, and then employ an unstructured hierarchy to organize the hierarchical framework. For instance, in learning arithmetic operations, addition serves as the foundation skill for the other three operations, suggesting mastery of addition enables mastery of any of the subsequent attributes (subtraction, multiplication, or division). Furthermore, determining the optimal number of basic skills to be identified when using a minimally restricted hierarchy to enhance classification efficiency represents an important avenue for future investigation.

This study's limitations include the restricted scope of manipulations and the lack of real-world data. Specifically, only limited factors and levels were considered in the simulation studies. Hence, future studies can explore a broader range of factors or levels (e.g., more CDMs, Q-matrix structures, attributes, or complexity in the combinations of heterogeneity). This would give a better picture of the effects and make the findings more useful.

References

Crowley, M. L. (1987). The van Hiele model of the development of geometric thought. *Learning and teaching geometry*, *K-12*, *1*, 1-16.

de La Torre, J. (2009). DINA model and parameter estimation: A didactic. Journal of *Educational and Behavioral Statistics*, *34*(1), 115-130. https://doi.org/10.3102/1076998607309474

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

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*(3), 205–237. http://www.jstor.org/stable/1435314

Liu, R. (2018). Misspecification of attribute structure in diagnostic measurement. *Educational and Psychological Measurement*, *78*(4), 605-634. https://doi.org/10.1177/0013164417702458

Liu, R., & Huggins-Manley, A. C. (2016). The specification of attribute structures and ist effects on classification accuracy in diagnostic test design. In L. A. van der Ark, D. M. Bolt, W.-C. Wang, J. A. Douglas, & M. Wiberg (Eds.), *Quantitative psychology research* (pp. 243-254). New York, NY: Springer. https://doi.org/10.1007/978-3-319 38759-8_18

Mayer, R. E. (1996). Learners as information processors: Legacies and limitations of educational psychology's second. *Educational Psychologist*, *31*(3–4), 151–161. https://doi.org/10.1080/00461520.1996.9653263

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/

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

Tatsuoka, K. K. (1983). Rule Space: An Approach for Dealing with Misconceptions Based on Item Response Theory. *Journal of Educational Measurement*, *20*(4), 345–354. http://www.jstor.org/stable/1434951

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. https://doi.org/10.1177/0146621607300286