

Assessing the Dimensionality of the Latent Attribute Space in Cognitive Diagnosis Through Testing for Conditional Independence



Youn Seon Lim and Fritz Drasgow

Abstract Cognitive diagnosis seeks to assess an examinee’s mastery of a set of cognitive skills called (latent) attributes. The entire set of attributes characterizing a particular ability domain is often referred to as the latent attribute space. The correct specification of the latent attribute space is essential in cognitive diagnosis because misspecifications of the latent attribute space result in inaccurate parameter estimates, and ultimately, in the incorrect assessment of examinees’ ability. Misspecifications of the latent attribute space typically lead to violations of conditional independence. In this article, the Mantel-Haenszel statistic (Lim & Drasgow in *J Classif*, 2019) is implemented to detect possible misspecifications of the latent attribute space by checking for conditional independence of the items of a test with parametric cognitive diagnosis models. The performance of the Mantel-Haenszel statistic is evaluated in simulation studies based on its Type-I-error rate and power.

Keywords Cognitive diagnosis model · Dimensionality · Mantel-haenszel statistic

1 Introduction

Cognitive diagnosis models (CDMs) try to account for the dependence among observations by latent dimensions that are related to the mastery or possession of cognitive skills, or “attributes” required for a correct response to an item. These models have received considerable attention in educational research because tests based on CDMs

Y. S. Lim (✉)

Department of Science Education,
Donald and Barbara Zucker School of Medicine at Hofstra/Northwell,
500 Hofstra University, Hempstead, NY 11549, USA
e-mail: younseon.lim@gmail.com; Younseon.Lim@hofstra.edu

F. Drasgow

School of Labor & Employment Relations, Department of Psychology, 504 E. Armory Avenue,
Champaign, IL 61820, USA

University of Illinois at Urbana-Champaign, 603 E. Daniel Street, Champaign, IL 61820, USA
e-mail: fdrasgow@illinois.edu

© Springer Nature Switzerland AG 2019

M. Wiberg et al. (eds.), *Quantitative Psychology*, Springer Proceedings
in Mathematics & Statistics 265, https://doi.org/10.1007/978-3-030-01310-3_17

promise to provide more diagnostic information about an examinee's ability than tests that are based on Item Response Theory (IRT) (Rupp et al., 2010). Specifically, whereas IRT defines ability as a unidimensional continuous construct, CDMs describe ability as a composite of K discrete, binary latent skill variables called attributes that define 2^K distinct classes of proficiency.

Like with other measurement models in assessment, the validity of a CDM depends on whether the latent attributes characterizing each proficiency class entirely determine an examinee's test performance, so that item responses can be assumed to be independent after controlling for the effect of the attributes. (This property of conditional independence is often called "local independence" in the IRT literature.) As Lord and Novick (1968) pointed out, the misspecification of the latent ability space underlying a test usually leads to violations of the conditional independence assumption that, in turn, result in inaccurate estimates of the model parameters and, ultimately, incorrect assessments of examinees' ability. For cognitive diagnosis, the assumption of conditional independence is equivalent to the assumption that the K attributes span the complete latent space. More to the point, violations of conditional independence are likely to occur if the latent attribute space has been misspecified in either including too few or too many latent attributes in the model.

Within the context of IRT models, various methods have been proposed for examining the dimensionality of the latent ability space underlying a test through checking for possible violations of conditional independence. Stout (1987), for example, developed DIMTEST, a nonparametric procedure for establishing unidimensionality of the test items through testing for conditional independence. Another instance is Rosenbaum's (1984) use of the Mantel-Haenszel statistic for assessing the unidimensionality of dichotomous items.

Lim and Drasgow (2019) proposed a nonparametric procedure for detecting misspecifications of the latent attribute space in cognitive diagnosis, which relies on the Mantel-Haenszel statistic to check for violations of conditional independence in the context of nonparametric cognitive diagnosis method approaches. This study extends the study of Lim and Drasgow (2019) by using the proposed statistic with parametric cognitive models for the estimation of proficiency classes.

2 The Mantel-Haenszel Test

Lim and Drasgow (2019) propose to use the Mantel-Haenszel (MH) chi-square statistic to test for the (conditional) independence of two dichotomous variables j and j' by forming the 2-by-2 contingency tables in conditioning on the levels of the stratification variable C . In their study, the stratification variable C is defined in terms of the latent attribute vector $\alpha_c = (\alpha_{c1}, \alpha_{c2}, \dots, \alpha_{cK})'$, for $c = 1, 2, \dots, 2^K$; that is, the different strata of C are formed by the 2^K proficiency classes.

Let $\{i_{j,j'c}\}$ denote the frequencies of examinees in the $2 \times 2 \times C$ contingency table. The marginal frequencies are the row totals $\{i_{1+c}\}$ and the column totals $\{i_{+1c}\}$, and i_{++c} represents the total sample size in the c th stratum. Then, the MH statistic

is defined as

$$MH\chi^2 = \frac{[\sum_c (i_{11c} - \sum_c E(i_{11c}))]^2}{\sum_c \text{var}(i_{11c})}, \tag{1}$$

where $E(i_{11c}) = i_{1+c}i_{+1c}/i_{++c}$ and $\text{var}(i_{11c}) = i_{0+c}i_{1+c}i_{+0c}i_{+1c}/i_{++c}^2(i_{++c} - 1)$. The stratum having minimum total sample size i_{++c} equal or larger than 1 is included. Under the null hypothesis of conditional independence of the items j and j' , for cognitive diagnosis models, the MH statistic has approximately a chi-square distribution with degrees of freedom equal to 1 if examinee's true latent attribute vectors are used as the levels of the stratification variable C . Assume that the odds ratio between j and j' is constant across all strata. Then the null hypothesis of independence is equivalent to an odds ratio of one

$$\text{Odds Ratio}_{MHj,j'} = \frac{1}{C} \sum_{c=1}^C \text{or}_{j,j'c}, \tag{2}$$

where $\text{or}_{j,j'c} = (i_{11c}i_{00c})/(i_{10c}i_{01c})$.

3 Simulation Studies

The finite test-length and sample-size properties of $MH\chi^2$ have been investigated in simulation studies. For each condition, item response data of sample sizes $I = 500$, or 2000 were drawn from a discretized multivariate normal distribution $MVN(0_K, \Sigma)$, where the covariance matrix Σ has unit variance and common correlation $\rho = 0.3$ or 0.6. The K -dimensional continuous vectors $\theta_i = (\theta_{i1}, \theta_{i2}, \dots, \theta_{iK})'$ were dichotomized according to

$$\alpha_{ik} = \begin{cases} 1, & \text{if } \theta_{ik} \geq \Phi^{-1} \frac{k}{K+1}; \\ 0, & \text{otherwise} \end{cases}$$

Test lengths $J = 20$ or 40 were studied with attribute vectors of length $K = 3$ or 5. The correctly specified Q -matrix for $J = 20$ is presented in Table 1 (Attributes with \star were used for Q -matrix ($K = 3$); attributes with $\star\star$ for Items 4 and 5). The Q -matrix for $J = 40$ was obtained by duplicating this matrix two times.

Data were generated from three different models: the DINA model, the additive-cognitive diagnosis model (A-CDM), and a saturated model (i.e., the generalized-DINA (G-DINA) model). For the DINA model, item parameters were drawn from Uniform (0, 0.3). For the A-CDM and the saturated model, like Chen et al. (2013), the parameters were restricted as $P(\alpha_{ij}^*)_{\min} = 0.10$ and $P(\alpha_{ij}^*)_{\max} = 0.90$, where α_{ij}^* was the reduced attribute vector whose components are the required attributes for the j -th item (see de la Torre, 2011, more details). The R was used for the estimation in this study (e.g., Robitzsch, Kiefer, George, & Uenlue, 2015) in which model parameter estimation was performed by maximization of the marginal likelihood.

Table 1 Correctly specified Q ($K = 5$)

Item	k_1^*	k_2^*	k_3^*	k_4	k_5	Item	k_1^*	k_2^*	k_3^*	k_4	k_5
1	1	0	0	0	0	11	1	1	0	0	0
2	0	1	0	0	0	12	1	1	0	0	1
3	0	0	1	0	0	13	1	0	0	1	1
4	0**	0	0	1**	0**	14	0	1	0	1	1
5	0**	0	0	0**	1**	15	0	0	1	1	0
6	1	1	0	1	1	16	1	0	0	1	0
7	1	0	0	1	0	17	0	1	0	0	1
8	0	1	0	1	0	18	0	0	1	0	0
9	0	0	1	1	0	19	1	0	0	1	0
10	0	1	0	1	0	20	1	0	0	1	1

For each condition, a set of item response vectors was simulated for 100 replications. The proposed MH statistic, Chi-squared statistic $x_{jj'}$ (Chen and Thissen, 1997), absolute deviations of observed and predicted corrections $r_{jj'}$ (Chen et al. 2013), and their corresponding p-values were computed for all $(J \times (J - 1))/2$ item-pairs in an individual replication.

4 Results

Across 100 trials for each condition, the proportion of times the p-value of each item-pair was smaller than the significance level 0.05 was recorded and is summarized in the tables shown below.

Type I Error Study In this simulation study, the correctly specified Q -matrices ($K = 5$, or $K = 3$) were used to fit the data to examine type I error rates. Table 2 shows that most type I error rates of the three different statistics were around the nominal significance level 0.05. The Chi-squared test statistic $x_{jj'}$ was conservative, with type I error rates below 0.024. The MH statistic got consistent under all conditions when item $J = 40$, confirming the asymptotic consistency. In the condition of $K = 5$, $J = 20$, and $I = 2000$, the type I error rates of the MH test slightly increased over the nominal rate in the A-CDM and the saturated model for the difficulty of correct classification.

Power Study: 20% misspecified Q -matrix For each replication, 20% of q_{jk} entries of the correctly specified Q -matrices ($K = 5$, or $K = 3$) were randomly misspecified. It is over-specification when q -entries of 0 are incorrectly coded as 1, and it is underspecification when q -entries of 1 are incorrectly coded as 0. Table 3 shows that the average rejection rates of all $J \times (J - 1) \times 1/2$ item pairs result in relatively low in the MH test (i.e., 0.310 or below in Non-Parametric Model, 373 or below in

Table 2 Type I error study

		$J = 20$						$J = 40$					
K	I	α with $\rho = 0.3$			α with $\rho = 0.6$			α with $\rho = 0.3$			α with $\rho = 0.6$		
		MH	x_{ij}'	r_{ij}'	MH	x_{ij}'	r_{ij}'	MH	x_{ij}'	r_{ij}'	MH	x_{ij}'	r_{ij}'
<i>DINA model</i>													
3	500	0.042	0.019	0.044	0.045	0.014	0.033	0.048	0.017	0.053	0.042	0.020	0.053
	2000	0.046	0.023	0.052	0.045	0.015	0.033	0.049	0.019	0.052	0.048	0.019	0.045
5	500	0.040	0.016	0.042	0.040	0.010	0.035	0.047	0.017	0.050	0.049	0.011	0.039
	2000	0.039	0.011	0.033	0.053	0.008	0.027	0.046	0.024	0.059	0.049	0.014	0.043
<i>A-CDM</i>													
3	500	0.036	0.009	0.029	0.031	0.009	0.026	0.039	0.011	0.030	0.036	0.011	0.028
	2000	0.048	0.013	0.030	0.049	0.010	0.026	0.048	0.010	0.029	0.047	0.010	0.026
5	500	0.021	0.020	0.038	0.024	0.015	0.033	0.039	0.018	0.046	0.034	0.014	0.039
	2000	0.073	0.016	0.036	0.065	0.011	0.031	0.040	0.019	0.047	0.043	0.014	0.039
<i>Saturated model</i>													
3	500	0.034	0.010	0.025	0.033	0.009	0.026	0.040	0.010	0.029	0.035	0.011	0.028
	2000	0.047	0.010	0.028	0.045	0.010	0.025	0.046	0.010	0.029	0.047	0.009	0.026
5	500	0.024	0.014	0.036	0.020	0.013	0.031	0.041	0.018	0.047	0.035	0.015	0.041
	2000	0.067	0.012	0.034	0.057	0.010	0.030	0.041	0.018	0.045	0.043	0.014	0.039

Table 3 Power study: 20% misspecified Q .

K		I		$J = 40$															
				α with $\rho = 0.3$				α with $\rho = 0.6$				α with $\rho = 0.3$				α with $\rho = 0.6$			
				MH	x_{ij}'	r_{ij}'	MH	x_{ij}'	r_{ij}'	MH	x_{ij}'	r_{ij}'	MH	x_{ij}'	r_{ij}'	MH	x_{ij}'	r_{ij}'	
<i>DINA model</i>																			
3	500	0.132	0.250	0.305	0.117	0.210	0.271	0.111	0.266	0.322	0.100	0.238	0.296						
	2000	0.235	0.400	0.449	0.209	0.393	0.472	0.143	0.397	0.438	0.134	0.379	0.430						
5	500	0.161	0.199	0.275	0.143	0.168	0.239	0.111	0.257	0.330	0.124	0.242	0.326						
	2000	0.373	0.465	0.536	0.294	0.360	0.448	0.105	0.186	0.263	0.207	0.394	0.470						
<i>A-CDM</i>																			
3	500	0.164	0.352	0.390	0.165	0.362	0.399	0.105	0.341	0.373	0.113	0.366	0.396						
	2000	0.241	0.478	0.510	0.258	0.456	0.491	0.141	0.423	0.454	0.156	0.434	0.463						
5	500	0.121	0.149	0.182	0.074	0.131	0.181	0.047	0.200	0.249	0.049	0.174	0.232						
	2000	0.149	0.196	0.230	0.169	0.270	0.338	0.106	0.347	0.389	0.124	0.347	0.387						
<i>Saturated model</i>																			
3	500	0.118	0.342	0.388	0.110	0.337	0.372	0.101	0.336	0.379	0.090	0.348	0.384						
	2000	0.206	0.470	0.493	0.270	0.578	0.614	0.131	0.441	0.473	0.130	0.459	0.490						
5	500	0.062	0.139	0.189	0.056	0.110	0.164	0.048	0.187	0.238	0.041	0.162	0.216						
	2000	0.152	0.285	0.343	0.139	0.230	0.287	0.102	0.301	0.349	0.081	0.279	0.333						

DINA model, 0.258 or below in A-CDM, 0.270 or below in saturated model). When $K = 5$, and $I = 500$, the power rates appear to be low (i.e., 0.074 or below) in the A-CDM, and the saturated model. They are rather complex models. It is very likely for small sample size to increase the difficulty of accurate model estimation.

Power Study: Over-specified Q -matrix For each replication, a data set was generated with the Q -matrix ($K = 3$) that is embedded as a subset of the Q -matrix ($K = 5$) in Table 1. The data was fitted with the Q -matrix ($K = 5$) to over-specify the correctly specified Q -matrix ($K = 3$). A dimension (total 9 items) or two dimensions (total 4 items) were over-specified. The results were consistent with what Chen et al. (2013) found.

As Table 4 shows, all statistics were insensitive to over-specified Q -matrices when the true models were the saturated model or the A-CDM. The average power rates of the item pairs where both items were over-specified in the same dimension were Non-Parametric Model = 0.074, MH = 0.052, $x_{jj'}$ = 0.181, and $r_{jj'}$ = 0.220, and those of the item pairs where either item was over-specified were MH = 0.058, $x_{jj'}$ = 0.104, and $r_{jj'}$ = 0.137 when the true model was the DINA model. If more attributes are included in the Q -matrix than required, as Rupp et al. (2010) indicated, conditional independence may still be preserved, because true attribute vector may be embedded in subcomponents of the modeled vector, resulting in a model that is too complex but preserves conditional independence. This finding implies that unlike the other statistics, the MH statistic is inappropriate to be used for the detection of the over-specified Q -matrices when the true model is the DINA model.

Power Study: Under-specified Q -matrix A data set was generated with the Q -matrix ($K = 5$) in Table 1. The data was fitted with the embedded Q -matrix ($K = 3$) in each replication. A dimension (total 9 items) or two dimensions (total 4 items) were under-specified. The average power rates of the item pairs where both items were under-specified in the same dimension were MH = 0.572, $x_{jj'}$ = 0.669, and $r_{jj'}$ = 0.735, with power relatively consistent across all conditions as shown in Table 5. The average rejection rates across item pairs where either item was under-specified were MH = 0.124, $x_{jj'}$ = 0.144, and $r_{jj'}$ = 0.201. The power rates slightly increased when $J = 40$, $I = 2000$, or the true model is the A-CDM in all statistics. Taking this finding into account, like the other statistics, the MH test is sensitive to Q -underspecification and has high power in all conditions.

Power Study: Model Misspecification In this simulation study, a correctly specified Q -matrix ($K = 3$ or 5) was used, but with a misspecified cognitive diagnosis models. As Chen et al. (2013) indicated, no statistics detected the model misspecification in all conditions when the fitted model was the saturated model, and the true models were the DINA model and the A-CDM (i.e., 0.052 or below for MH, 0.024 or below for $x_{jj'}$, and 0.059 or below for $r_{jj'}$). Due to limited space, the output is not included. The results in Table 6 show that the rejection rates of the MH statistic were low (i.e.,

Table 4 Power study: Over-specified Q with true $K = 5$.

I	S	$J = 40$															
		α with $\rho = 0.3$				α with $\rho = 0.6$				α with $\rho = 0.3$				α with $\rho = 0.6$			
		MH	x_{ij}'	r_{ij}'	r_{ij}'	MH	x_{ij}'	r_{ij}'	r_{ij}'	MH	x_{ij}'	r_{ij}'	r_{ij}'	MH	x_{ij}'	r_{ij}'	r_{ij}'
<i>DINA model</i>																	
500	O	0.038	0.167	0.218	0.043	0.150	0.191	0.048	0.145	0.179	0.048	0.137	0.168				
	E	0.059	0.095	0.128	0.046	0.080	0.118	0.051	0.086	0.114	0.049	0.086	0.115				
2000	O	0.073	0.271	0.311	0.063	0.249	0.295	0.050	0.181	0.221	0.048	0.150	0.177				
	E	0.075	0.139	0.175	0.079	0.156	0.194	0.048	0.100	0.134	0.056	0.094	0.119				
<i>A-CDM</i>																	
500	O	0.031	0.008	0.021	0.032	0.009	0.018	0.037	0.006	0.021	0.033	0.003	0.016				
	E	0.047	0.010	0.024	0.040	0.008	0.024	0.040	0.006	0.023	0.039	0.006	0.024				
2000	O	0.039	0.012	0.020	0.038	0.019	0.025	0.045	0.004	0.018	0.044	0.003	0.015				
	E	0.040	0.011	0.024	0.043	0.013	0.025	0.046	0.006	0.023	0.046	0.006	0.022				
<i>Saturated model</i>																	
500	O	0.035	0.007	0.020	0.038	0.004	0.016	0.038	0.007	0.022	0.033	0.003	0.017				
	E	0.052	0.006	0.020	0.049	0.004	0.018	0.040	0.007	0.027	0.039	0.005	0.023				
2000	O	0.041	0.012	0.019	0.041	0.085	0.096	0.044	0.006	0.021	0.042	0.005	0.017				
	E	0.042	0.018	0.020	0.041	0.045	0.058	0.042	0.007	0.024	0.046	0.007	0.023				

Table 5 Power study: under-specified Q with true $K=6$

I	S	$J = 40$															
		α with $\rho = 0.3$				α with $\rho = 0.6$				α with $\rho = 0.3$				α with $\rho = 0.6$			
		MH	x_{ij}'	r_{ij}'		MH	x_{ij}'	r_{ij}'		MH	x_{ij}'	r_{ij}'		MH	x_{ij}'	r_{ij}'	
<i>DINA model</i>																	
500	U	0.390	0.462	0.541	0.672	0.595	0.725	0.488	0.550	0.608	0.319	0.415	0.499				
	E	0.093	0.088	0.149	0.150	0.064	0.136	0.145	0.130	0.194	0.094	0.129	0.194				
2000	U	0.532	0.608	0.648	0.489	0.599	0.672	0.538	0.624	0.658	0.780	0.845	0.900				
	E	0.210	0.245	0.314	0.106	0.171	0.225	0.290	0.285	0.359	0.189	0.180	0.259				
<i>A-CDM</i>																	
500	U	0.464	0.553	0.637	0.453	0.581	0.675	0.551	0.642	0.722	0.535	0.682	0.766				
	E	0.093	0.061	0.100	0.064	0.081	0.124	0.051	0.092	0.138	0.045	0.139	0.196				
2000	U	0.744	0.846	0.890	0.749	0.866	0.911	0.860	0.913	0.941	0.859	0.928	0.951				
	E	0.275	0.139	0.190	0.184	0.208	0.259	0.096	0.206	0.264	0.077	0.313	0.363				
<i>Saturated model</i>																	
500	U	0.371	0.474	0.565	0.344	0.485	0.579	0.453	0.548	0.641	0.409	0.557	0.652				
	E	0.074	0.040	0.081	0.054	0.061	0.103	0.055	0.058	0.104	0.045	0.089	0.139				
2000	U	0.633	0.778	0.831	0.611	0.776	0.826	0.764	0.862	0.902	0.714	0.856	0.899				
	E	0.241	0.118	0.174	0.147	0.173	0.231	0.110	0.151	0.216	0.080	0.241	0.304				

Table 6 Power study: model misspecification

K	I	J = 20						J = 40					
		α with $\rho = 0.3$			α with $\rho = 0.6$			α with $\rho = 0.3$			α with $\rho = 0.6$		
		MH	x_{ij}'	r_{ij}'	MH	x_{ij}'	r_{ij}'	MH	x_{ij}'	r_{ij}'	MH	x_{ij}'	r_{ij}'
<i>True = DINA model, Fitted = A-CDM</i>													
3	500	0.113	0.146	0.135	0.081	0.129	0.090	0.075	0.308	0.192	0.071	0.250	0.173
	2000	0.100	0.391	0.310	0.115	0.453	0.383	0.089	0.429	0.236	0.075	0.144	0.134
5	500	0.134	0.279	0.269	0.080	0.262	0.234	0.101	0.620	0.376	0.093	0.368	0.308
	2000	0.296	0.632	0.493	0.186	0.512	0.366	0.114	0.778	0.603	0.145	0.930	0.450
<i>True = A-CDM, Fitted = DINA model</i>													
3	500	0.113	0.146	0.135	0.081	0.129	0.090	0.075	0.308	0.192	0.071	0.250	0.173
	2000	0.100	0.391	0.310	0.115	0.453	0.383	0.089	0.429	0.236	0.075	0.144	0.134
5	500	0.134	0.279	0.269	0.080	0.262	0.234	0.101	0.620	0.376	0.093	0.368	0.308
	2000	0.296	0.632	0.493	0.186	0.512	0.366	0.114	0.778	0.603	0.145	0.930	0.450

0.186 or below with few exceptions when the true model was the DINA model, and the fitted model was the A-CDM, 0.097 or below with few exceptions when verse versa). When the true model was the A-CDM, and the fitted model was the DINA model, the power rates were even lower because the DINA model is simpler than the A-CDM.

5 Discussion

A Mantel-Haenszel(MH) statistic proposed by Lim and Drasgow (2019) was evaluated for detecting misspecifications of the latent attribute space in parametric cognitive diagnosis models; that is, the Q -matrix might contain too many or too few latent attributes. (Recall that a misspecified latent attribute space may result in inaccurate parameter estimates that will cause incorrect assessments of examinees' ability.) The proposed MH statistic uses as the levels of the stratification variable the different proficiency classes, with examinees' individual attribute vectors—that identify proficiency class membership—estimated from the data. Simulation studies were conducted for investigating the diagnostic sensitivity of the MH statistic in terms of Type-I-Error rate and power under a variety of testing conditions. Across different sample sizes, test lengths, number of attributes defining the true attribute space, and levels of correlation between the attributes, the MH statistic consistently attained a Type-I-Error rate that was typically close to the nominal $0.05 - \alpha$ -level when the data were generated using the true Q -matrix based on the correctly specified latent attribute space. When the data were generated using a Q -matrix based on an under-specified latent attribute space, the MH statistic displayed moderate power in detecting the resulting conditional dependence among test items. In summary, the MH statistic might be a promising tool for uncovering possible misspecifications of the latent attribute space in cognitive diagnosis. Further research is needed to investigate the specific factors that affect the power of the MH statistic; especially, when the latent attribute space has been over-specified (i.e., too many attributes have been included).

References

- Chen, J., de la Torre, J., & Zhang, Z. (2013). Relative and absolute fit evaluation in cognitive diagnosis modeling. *Journal of Educational Measurement*, 50, 123–140.
- Chen, W., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, 22, 265–289.
- de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, 76, 179–199.
- Lim, Y. S., & Drasgow, F. (2019). Conditional independence and dimensionality of nonparametric cognitive diagnostic models: A test for model fit. *Journal of Classification*.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

- Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of National Cancer Institute*, 22, 719–748.
- Robitzsch, A., Kiefer, T., George, A. C., & Uenlue, A. (2015). CDM: Cognitive diagnostic modeling. R package version 3.4–21.
- Rosenbaum, P. R. (1984). Testing the conditional independence and monotonicity assumption of item response theory. *Psychometrika*, 49, 425–436.
- Rupp, A., Templin, J., & Henson, R. (2010). *Diagnostic assessment: Theory, methods, and applications*. New York: Guilford.
- Stout, W. (1987). A nonparametric approach for assessing latent trait dimensionality. *Psychometrika*, 52, 589–618.