The Effect of the Item–Attribute Relation on the DINA Model Estimations in the Presence of Missing Data[*]

Kayıp Veri Varlığında DINA Model Madde-Özellik İlişkisinin Parametre Kestirimine Etkisi

Ömür Kaya KALKAN[**], Tahsin Oğuz BAŞOKÇU[***]

Abstract

The objective of this study is to investigate the relation between the number of items and attributes and to analyze the manner in which the different rates of missing data affect the model estimations based on the simulation data. A Q-matrix contains 24 items, and data are generated using four attributes. A dataset of $n = 3,000$ is generated by associating the first, middle, and final eight items in the Q-matrix with one, two, and three attributes, respectively, and 5%, 10%, and 15% of the data have been randomly deleted from the first, middle, and final eight-item blocks in the Q-matrix, respectively. Subsequently, imputation was performed using the multiple imputation (MI) method with these datasets, 100 replication was performed for each condition. The values obtained from these datasets were compared with the values obtained from the full dataset. Thus, it can be observed that an increase in the amount of missing data negatively affects the consistency of the DINA parameters and the latent class estimations. Further, the latent class consistency becomes less affected by the missing data as the number of attributes associated with the items increase. With an increase in the number of attributes associated with the items, the missing data in these items affect the consistency level of the *g* parameter (guess) less and the *s* parameter (slip) more. Furthermore, it can be observed from the results that the test developers using the cognitive diagnosis models should specifically consider the item–attribute relation in items with missing data.

Keywords: DINA model, missing data, latent class estimates, item–attribute relation.

Cited:

Kalkan, Ö.K., & Başokçu, T.O. (2019). The effect of the item–attribute relation on the DINA model estimations in the presence of missing data. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi, 46*, 290-306. doi: 10.9779/pauefd.546797

** Asst. Prof., Pamukkale University, Faculty of Education, Department of Educational Sciences, Division of Educational Measurement and Evaluation, Denizli. ORCID : 0000-0002-4821-0045
e-mail: kayakalkan@pau.edu.tr
*** Assoc. Prof., Ege University, Faculty of Education, Department of Educational Sciences, Division of Educational Measurement and Evaluation, Izmir. ORCID: 0000-0001-7088-4268
e-mail: tahsin.oguz.basokcu@ege.edu.tr

Öz

Bu araştırmanın amacı, farklı oranlarda kayıp veri varlığında madde-özellik sayısı ilişkisinin, DINA model kestirimlerini nasıl etkilediğini simülasyon verileri üzerinden incelemektir. Verilerin üretilmesinde dört özellik ve 24 maddeden oluşan bir Q matris kullanılmıştır. Q matrixteki ilk, orta ve son 8 madde sırasıyla 1, 2 ve 3 özellikle ilişkilendirilerek 3,000 kişilik bir veri seti üretilmiş ve bu verilerde yer alan her 8 maddelik bloktan sıras ile %5, %10 ve %15 veri rassal silinmiştir. Ardından, bu veri setlerine M yöntemi ile imputasyon yapılmıştır. Bu işlemler, her bir koşul için 100 kez tekrarlanmıştır. Bu veri setlerinden elde edilen kestirimler, kayıpsız veri setinden elde edilen değerler ile karşılaştırılmıştır. Araştırmanın bulgular, kayıp veri miktarındaki artışın, DINA model parametre ve örtük sınıf kestirimlerindeki tutarlılığı, olumsuz yönde etkilediğini göstermiştir. Maddenin ilişkili olduğu özellik sayısı arttıkça örtük sınıf uyumu kayıp veriden daha az etkilenmiştir. Maddenin ilişkili olduğu attribute sayısı arttıkça bu maddelerde gözlenen kayıp veri, testin g parametresi uyum düzeyini daha az, s parametresini daha çok etkilemiştir. Araştırmanın sonuçlar, özellikle CDM modellerini kullanan test geliştiricilerinin kayıp veri gözlenen maddelerde, madde-özellik ilişkisini göz önünde bulundurmaları gerektiğini göstermektedir.

Anahtar sözcükler: DINA model, kayıp veri, örtük sınıf kestirimi, madde-özellik ilişkisi.

Atıf:

Kalkan, Ö.K., ve Başokçu, T.O. (2019). Kayıp veri varlığında DINA model madde-özellik ilişkisinin parametre kestirimine etkisi. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi, 46*, 290-306. doi: 10.9779/pauefd.546797

Introduction

Recently, there has been a considerable increase in interest in cognitive diagnosis models (CDMs) with respect to educational measurement. The item response theory (IRT) and the classical test theory attempted to assess the overall ability of the respondents, whereas the CDMs focused on diagnosing the weaknesses and strengths of the examinees based on a set of specific attributes. Thus, CDMs provide considerably detailed and fine-grained assessments examining the execution of the examinees with respect to specific skills instead of simply reporting a single overall test score (Sen and Bradshaw, 2017). Further, the diagnostic information that is obtained based on the CDM assessment can be used for accurately measuring the learning status of students and for facilitating better instruction and intervention (Chen, 2017; Sorrel et al., 2016). CDMs are considered to be multidimensional, multivariate, discrete, and latent trait models (de la Torre & Lee, 2010; Sorrel et al., 2016), and different theoretical frameworks, such as the restricted latent class models (Haertel, 1989), the IRT (Embretson & Reise, 2000; Fischer, 1973; van der Linden & Hambleton, 2013), and the rule space method (Tatsuoka, 1983), have contributed to the development of these models. On the basis of different approaches, can also be referred to as multiple classification models (Maris, 1999), restricted latent class models (Haertel, 1989), CDMs (Henson & Douglas, 2005), and structured IRT models (Rupp & Mislevy, 2007). Various CDMs draw attention in terms of the approaches that they use in parameter estimates and the item–attribute relation. Some of these approaches include deterministic inputs, noisy "AND" gate (DINA; Junker & Sijtsma, 2001), deterministic inputs, noisy "OR" gate (DINO; Templin & Henson, 2006), noisy inputs, deterministic "AND" gate (NIDA; Junker & Sijtsma, 2001), noisy input deterministic "OR" gate (NIDO; Templin & Henson, 2006), reparameterized unified model (RUM; Hartz, 2002), log-linear CDM (Henson, Templin & Willse, 2009), and general diagnostic model (GDM; von Davier, 2005). Despite the presence of several CDM formulations, the DINA model was selected in this study because of its simplicity, ease of interpretability, and good model–data fit (de la Torre & Lee, 2010). Further, a concise explanation of the DINA model is presented in the following section on the basis of the study conducted by Junker and Sijtsma (2001).

The DINA Model

The DINA model has recently become one of the most preferred CDMs. It is one of the simplest multiple classification models and is preferred by researchers because it provides clear interpretability and good model–data fit along with simplicity (de la Torre, & Douglas, 2008; de la Torre, & Lee, 2010; Rupp, & Templin, 2008). The complexity of the DINA model is unaffected by the number of properties specified in the Q-matrix, which is unlike that observed in the NIDA and RUM models, because parameters are estimated for each item and not for each attribute (Rupp, & Templin, 2008). In addition, the DINA model analyses are easily performed using packages (e.g., [CDM], Robitzsch, Kiefer, George, & Uenlue, 2018; [GDINA], Ma, & de la Torre, 2018) or using codes compatible with open source software such as R (R Core Team, 2018) and Ox (Doornik, 2018). Similar to other CDMs, the DINA model requires the construction of a Q-matrix by domain experts (Tatsuoka, 1983). Q-matrix is the key component of the test construction process based on CDM. Q-matrix comprises 1×0 in which the items are located in rows and the attributes that are required for an item to be accurately answered are located in columns. In terms of the conjunctive model, the DINA model assumes that a

candidate should possess all the necessary attributes specified in the Q-matrix to accurately answer an item. In the context of the restricted latent class, the latent response variables of the DINA model are presented as (Junker & Sijtsma, 2001)

$$\xi_{ij} = \prod_{k:Q_{jk}=1} \alpha_{ik} = \prod_{k=1}^{K} \alpha_{ik}{}^{Q_{jk}},$$

denoting whether an examinee possesses all the required attributes for accurately answering item j. The latent vectors $\alpha_{i.} = (\alpha_{i1}, \dots, \alpha_{iK})$ indicate the presence of skill k, whereas the latent response vectors (LRVs) $\xi_{i.} = (\xi_{i1}, \dots, \xi_{iJ})$ express the conjunctive process of the model. Thus, the LRV requires an examinee to possess all the predetermined skills for succeeding in a task. In the deterministic sense, the LRV should be identical to the observed response vector; however, because of the slip (s) and guess (g, noise) parameters, the LRV only represents the ideal response pattern (de la Torre, 2009). Further, the DINA model generates s and g parameters for each item. The item response function that accurately calculates the probability of an examinee answering an item is

$$P\left[X_{ij} = 1 \middle| \alpha, s, g\right] = (1 - s_j)^{\xi_{ij}} g_j^{1-\xi_{ij}} \equiv P_j(\alpha_{i.}).$$

It has been considered that each ξ_{ij} is the function of the õANDö gate, and the probability of an examinee accurately answering an item will be 0 or 1 when there is no g and s. By presuming local independency and independency among examinees, the joint likelihood for the DINA model can be obtained as

$$P\left[X_{ij} = x_{ij}, \forall i, j \middle| \alpha, s, g\right] = \prod_{i=1}^{N} \prod_{j=1}^{J} [(1 - s_j)^{x_{ij}} s_j^{1-x_{ij}}]^{\xi_{ij}} [g_j^{x_{ij}} (1 - g_j)^{1-x_{ij}}]^{1-\xi_{ij}}$$

Subsequently, the classification accuracy and item parameter estimates of the DINA model are affected by various factors, for example, the construction of the Q-matrix, the structure of the latent class, the characteristic of prior distribution, the sample size, the value of the g and s *para*meters, and the estimation procedure (de la Torre, Hong, & Deng, 2010). Additionally, the missing data can affect the accuracy and consistency of the parameter estimates and may lead to biased estimates in the latent class models (Winship, Mare, & Warren, 2002). Within the scope of this research, the missing data and methods for handling the missing data are briefly addressed, and interested readers are encouraged to refer to Little and Rubin (2002) for obtaining detailed information.

One of the most common problems encountered during educational and psychological research is missing data (Zhang & Walker, 2008). This problem occurs when examinees fail to answer the items because of several reasons such as the lack of information, shyness, reluctance, lack of time, or the design of the researchers in a planned way (Graham, Taylor, Olchowski, & Cumsille, 2006; Sijtsma & van der Ark, 2003; Little & Rubin, 2002). Consequently, missing data can lead to serious problems such as biased parameter estimates, information loss, decrease in statistical power, inflation of standard errors, and attenuation of the generalizability of the conclusions (Dong & Peng, 2013). The most important issues to be consider in regard to the

missing data are the rate, pattern, and mechanism of the missing data. Tabachnick and Fidell (2007) stated that 5% or lower rate of the missing data is negligible in large samples even though there is no exact criterion that has been provided in the literature. However, Bennet (2001) expressed that a missing data rate of more than 10% can be attributed to biased estimation. The missing data pattern indicates the observable values in the dataset and the missing values; further, the missing data mechanism expresses whether there is a relation between the missing data pattern and the values of the variables in the data matrix (Little & Rubin, 2002). Little and Rubin (2002) stated that the following three mechanisms can lead to missing data: missing completely at random (MCAR); missing at random; and not missing at random. The MCAR mechanism is based on the assumption that the probability of a missing variable is independent of the remaining measured variables (Enders, 2010). In this study, the MCAR missing data mechanism was selected. The missing data are handled using several methods such as listwise deletion, pairwise deletion, mean substitution, regression substitution, pattern-matching imputation, stochastic regression, expectation maximization, multiple imputation (MI), and full information maximum likelihood. Herein, the MI method is preferred because it provides more accurate and less biased standard errors in case of parameter estimates when compared to those provided using single imputation methods (Finch, 2008; Schlomer, Bauman, & Card, 2010). The MI method creates multiple datasets by performing a prespecified number of imputations (e.g., typically 5 or 10) to the missing data. Subsequently, each of the completed datasets is analyzed, and the average of the parameter estimates is estimated to produce a single set of results (Peugh & Enders, 2004). However, introducing other methods for handling the missing data is beyond the scope of this study.

Consequently, missing data are likely to occur while collecting data from individuals via educational and psychological tests. Missing data, similar to several other factors, can affect the item parameter and latent class estimates of the DINA model (Başokçu, Kalkan, & Öğretmen, 2016). Therefore, it seems important to examine the robustness of the item parameter and latent class estimates in the presence of missing data. Additionally, determining the effect imposed by the missing data on the item–attribute relation will provide noteworthy contributions to construct tests based on cognitive diagnosis.

Method

We designed this study to determine the deviation of the DINA model estimations from the true item parameters when a missing data imputation procedure is used and also to specify the effects of the number of attributes that the item is related to with respect to this deviation. The basic logic of the proposed design is to determine the manner in which an increase in the number of attributes measured by an item from among the items that exhibit missing data in the CDM-based tests changes the test parameters. Further, we manipulated both the level and type of missing data for the test in the pattern. We manipulated the rate of missing data and the number of attributes related to an item. Thus, in CDM analyses, we have attempted to reveal the manner in which the structure of the items in the observed missing data affects the model fit. The relation between an item and attribute is a prior definition in the CDM approach. This research aims to provide functional information for the test developers before testing rather than providing the posterior information such as the item difficulty or item discrimination.

Simulation Study

Design and Analysis

Herein, we examined the attributeóitem relations using simulation data. For this purpose, we used the õCDMö (Robitzsch, Kiefer, Uenlue, & Robitzsch, 2018) package of R (R Core Team, 2018) for data generation and parameter estimations based on CDM; we also used the prodNA function in the õmissForestö (Stekhoven, 2016) package of R for deleting the data completely random up to the amount specified in the dataset and the õmiceö (van Buuren, & Groothuis-Oudshoorn, 2011) package of R for performing MI. Table 1 presents the Q-matrix used for the generation of data from four attributes and 24 items.

Table 1. *Q-matrix*

	item	1	2	3	4		item	1	2	3	4		item	1	2	3	4
Block1-B1	1	1	0	0	0	*Block2-B2*	9	1	1	0	0	*Block3-B3*	17	1	1	1	0
	2	0	1	0	0		10	1	0	1	0		18	1	1	0	1
	3	0	0	1	0		11	1	0	0	1		19	1	0	1	1
	4	0	0	0	1		12	0	1	1	0		20	0	1	1	1
	5	1	0	0	0		13	0	1	0	1		21	1	1	1	0
	6	0	1	0	0		14	0	0	1	1		22	1	1	0	1
	7	0	0	1	0		15	0	1	0	1		23	1	0	1	1
	8	0	0	0	1		16	1	0	1	0		24	0	1	1	1

In this study, the number of attributes for data generation is four. Thus, there will be potentially 16 attribute patterns that may be available for participants and 15 property patterns that may be available for items. In this case, the minimum number of items should be 15 to ensure that an assessment reflects all the possible attribute patterns through items (Rupp & Templin, 2008). On the basis of the minimum number of items, a test was constructed using 24 items. In the Q-matrix configuration, the first eight items (1ó8, [block1, B1]) were associated with one attribute, the next eight items (9ó16, [block2, B2]) were associated with two attributes, and the final eight items (17ó24, [block3, B3]) were associated with three attributes. Thus, in each block (B1, B2, and B3) and, overall, we constructed a Q-matrix with an equal number of items associated with each attribute. Therefore, the total number of rows and columns of the Q-matrix for each block were equal (de la Torre, 2008, 2011; de la Torre ve Douglas, 2008; Rupp & Templin, 2008). Initially, a dataset with a sample size of 3,000 and in which *g* and *s* vary between 0.1 and 0.3 was generated. We obtained the *g* and *s* parameters and the latent class estimations by analyzing this dataset. We accepted these obtained parameters as the true parameters. Within the scope of this study, we determined three different missing data rates as 5%, 10%, and 15%. Furthermore, we randomly deleted 5%, 10%, and 15% of the data were randomly deleted from the initially produced complete dataset for the first 8 (B1), middle 8 (B2), and final 8 (B3) items, respectively. Subsequently, we performed imputation of these datasets using the MI method (*n* = 5). We performed 100 replications for each condition. Subsequently, we compared the *g* and *s* parameters and the latent class estimations obtained from these datasets with the true values obtained from the complete dataset. We performed these comparisons based on the absolute mean difference and the consistency percentages.

Findings

Latent Class Estimations

The imputations were performed using the MI method ($n = 5$) after 5%, 10%, and 15% of the data were randomly deleted from the B1, B2, and B3 blocks of the complete dataset. We performed 100 replications for each condition. The minimum, maximum, and mean values of the consistency percentage between the latent class estimations obtained from the complete datasets and the latent class estimation of the complete dataset were calculated. Figure 1 depicts these consistency percentages.
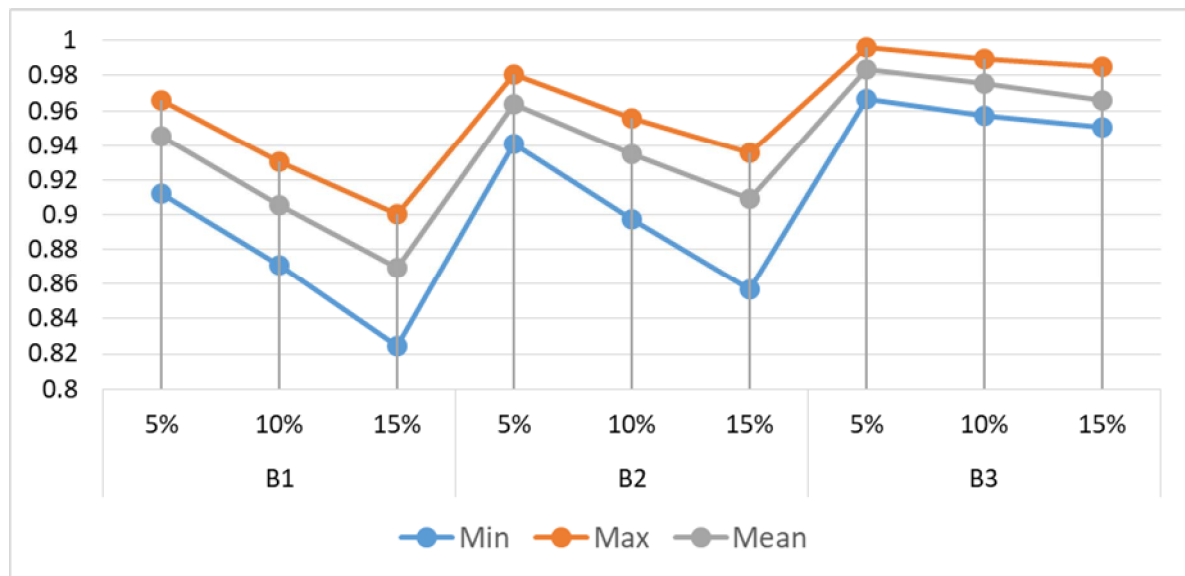

Figure 1. Consistency percentages of the latent class estimates

In Figure 1, the minimum values of the latent class estimations obtained from 100 different datasets for each condition are denoted in blue; further, the mean values are denoted in gray, whereas the maximum values are denoted in red. It can be observed from Figure 1 that the latent class estimations become approximately equal to the true value when the number of attributes associated with the items increases. For example, each item is related with one attribute in B1 block. After deleting 5% of the data from the B1 block and performing imputation using the MI method, the consistency of the obtained latent class estimations exhibited a minimum of 0.91, maximum of 0.97, and mean of 0.95. The values were observed to become 0.97, 1.00, and 0.96, respectively, in the B3 block, where each item was related with three attributes. Similar results were observed in case of other missing data values.

Furthermore, it was observed that an increase in the amount of missing data resulted in a decrease in the consistency values between the latent class estimates and the true values. However, the latent class estimations based on the B3 block were the least affected by the missing data, and they exhibited the most consistent estimations.

Item Parameters

The effect of the change in the percentage of missing data on the *g* parameter is depicted in Figure 2 by considering the number of attributes related to an item. In Figure 2, the B1 block is

gray, the B2 block is blue, and the B3 block is red. The RMSE values with respect to the g parameter obtained in case of deletion of data from the B1, B2, and B3 blocks are depicted in Figure 2 for each percentage amount. For example, the average RMSE value with respect to the g parameter was observed to be 0.005 when the amount of missing data for the B1 block, where the items were related with a single attribute, was 5%; the RMSE value with respect to the g parameter was observed to be approximately 0.007 when the percentage of missing data was 10% and 0.01 when the percentage of missing data was 15%. Further, a similar increase was observed in case of the B2 and B3 blocks. As the percentage of the missing data increased, the g parameters were observed to stray from their true values.
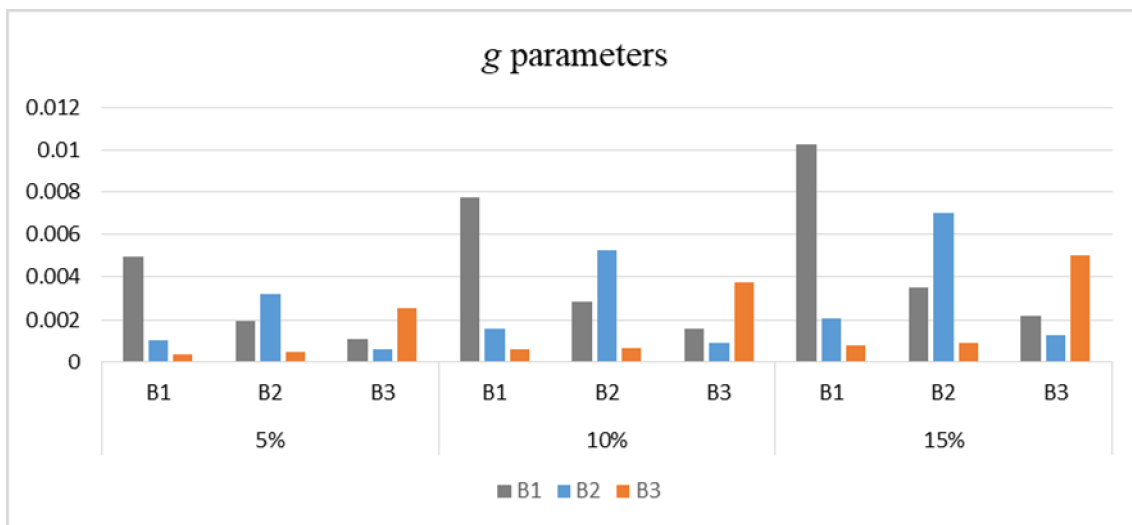

Figure 2. The itemó attribute relation and the effect of the percentage of missing data on the g parameters.

When the number of attributes related with the items was considered, the effects of the change in the percentage of missing data on the s parameter is depicted in Figure 3. In Figure 3, the B1 block is denoted in gray, the B2 block is denoted in blue, and the B3 block is denoted in red. Figure 3 denotes RMSE values obtained with respect to the s parameter when data are deleted from the B1, B2, and B3 blocks for each percentage amount. For example, the RMSE value with respect to the s parameter is 0.0048 when the amount of missing data for the B1 block, where items are related with a single attribute, was 5%; the RMSE value with respect to the s parameter was approximately 0.0073 when the percentage of missing data was 10% and 0.01 when the percentage of missing data was 15%. A similar increase was observed in case of the B2 and B3 blocks. As the percentage of the missing data increased, the s parameters were observed to stray from their true values.
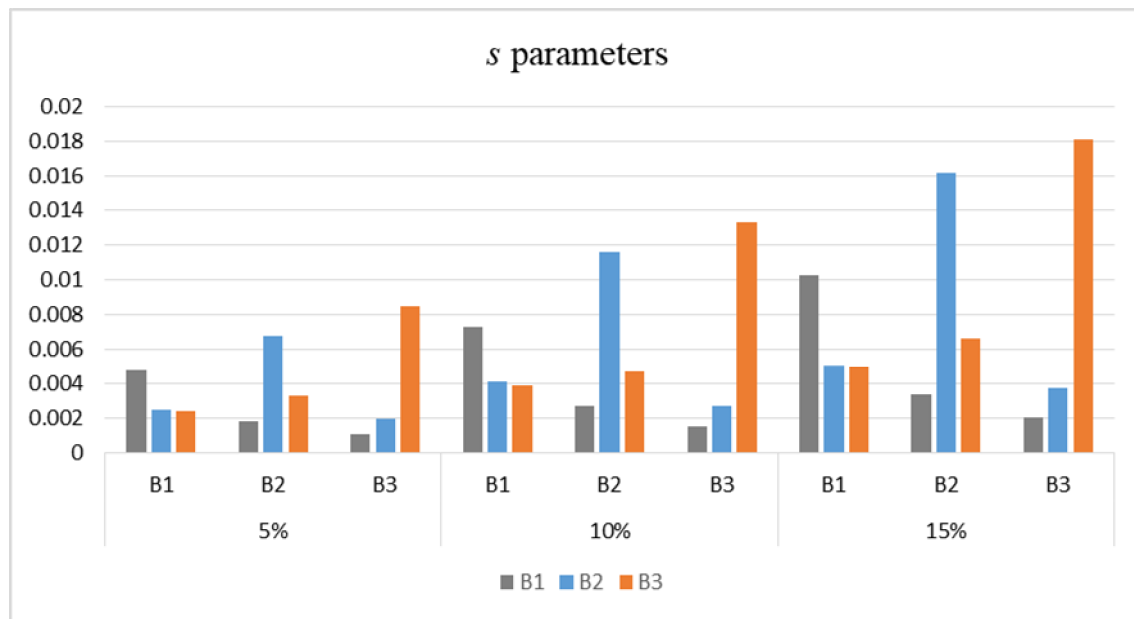
Figure 3. The item–attribute relation and the effect of the percentage of missing data on the *s* parameters.

In addition, when missing data were observed in items (B1) related with a single attribute, the RMSE values with respect to the *s* parameter for the other blocks in the test were almost equal. Furthermore, if the missing data were observed in the B2 and B3 blocks, the B1 block produced the closest values to the true value.

The general state of difference between the true values and the estimated values for the two parameters is presented in Table 2.

Table 2. The RMSE values for the true and estimated parameter differences

		g				*s*			
		B1	B2	B3	Mean	B1	B2	B3	Mean
5%	B1	0.0049	0.0020	0.0011	0.0027	0.0048	0.0018	0.0011	0.0026
	B2	0.0010	0.0032	0.0006	0.0016	0.0025	0.0068	0.0020	0.0038
	B3	0.0004	0.0005	0.0026	0.0011	0.0024	0.0033	0.0085	0.0047
	Mean	0.0021	0.0019	0.0014		0.0032	0.0040	0.0038	
10%	B1	0.0077	0.0029	0.0016	0.0041	0.0073	0.0028	0.0016	0.0039
	B2	0.0016	0.0052	0.0009	0.0026	0.0041	0.0116	0.0027	0.0062
	B3	0.0006	0.0007	0.0037	0.0017	0.0039	0.0047	0.0133	0.0073
	Mean	0.0033	0.0029	0.0021		0.0051	0.0064	0.0059	
15%	B1	0.0103	0.0035	0.0022	0.0053	0.0103	0.0034	0.0021	0.0052
	B2	0.0020	0.0070	0.0013	0.0035	0.0051	0.0161	0.0037	0.0083
	B3	0.0008	0.0009	0.0050	0.0023	0.0050	0.0066	0.0181	0.0099
	Mean	0.0044	0.0038	0.0028		0.0068	0.0087	0.0080	

Table 2 presents the RMSE values with respect to the manner in which the variance in the item block, where the missing data were observed, affected the *g* and *s* parameter estimations during the entire test. The average RMSE value with respect to the *g* parameters of the entire test was calculated to be 0.0021 if 5% of the missing data were from the B1 block, 0.0019 in the B2 block, and 0.0014 in the B3 block when the effect of variance in the item block, where the missing value observed, was analyzed by performing the *g* parameter

estimation of the entire test. Similar findings were also observed for 10% and 15% of missing data. Based on these observations, it can be stated that the missing data leads to more difference on g parameters of test than the items related with more attributes (B2 and B3) when the missing data is observed on the items related with a single attribute. In other words, the g parameter estimations obtained from the items related with a large number of attributes were observed to be less affected by the missing data when compared with the g parameter estimations obtained from the items related with less number of attributes.

In addition, the average RMSE value with respect to the s parameters of the entire test was calculated as 0.0032 if 5% of the missing data were located in the B1 block, 0.0040 in the B2 block, and 0.0038 in the B3 block when the effect of variance in the item block, where the missing value was observed, was analyzed with respect to the g parameter estimation during the entire test. Similar findings were observed for 10% and 15% of the missing values. Thus, it was observed that the missing data lead to less difference with respect to the s parameters of the test than the items related with more attributes (B2 and B3) when the missing data are observed on the items related with single attribute. In other words, the s-parameter estimations obtained from the items related to a single attribute were observed to be less affected by the missing data than the s-parameter estimations obtained from the items related with more attributes.

In conclusion, Table 2 presents that an increase in the missing data rate adversely affects the item parameters for all the conditions. Furthermore, it was observed that the values for a block exhibiting missing data for the item parameters denoted more deviation from the true value when compared to the deviation exhibited by the values of the remaining blocks. Therefore, when the average RMSE values for the entire test were considered in the presence of missing data, the g parameters were observed to be less affected by the missing data than the s *para*meters.

Discussion and Conclusion

This study discussed the manner in which the itemó attribute relation affected the item parameter and latent class estimates in case of different missing data rates. When the missing data rate was 5%, the g parameter RMSE values from the B1, B2, and B3 blocks were observed to vary between 0.0004 and 0.005; the s-parameter RMSE values were observed to change from 0.001 to 0.0085. The consistency mean of the latent class estimation varied between 0.95 and 0.98. This observation supports the finding reported by Tabachnick and Fidell (2007) in which 5% or less amount of missing data do not lead to significant problems in large samples. Increasing the amount of missing data was observed to increase the RMSE values with respect to both the g and s parameters, whereas the latent class estimations decreased the consistency values. This result regarding an increase in the missing data is, in fact, an expected situation. Increasing the amount of missing data was observed to negatively affect the consistency of the measurement results for all the models. This condition was observed both with respect to the change in item parameters and in the level of consistency of the latent classes.

Another finding was that the level of effect of the general parameters of the test was observed to be related to the items in which the missing data were observed. Because of the analysis, the parameters related to the items containing missing data varied from the true values at a significant level according to the parameters of the other items in the test, as expected. When the findings were examined, the parameter change of the block that contained the missing

data was higher than that observed in the other blocks for all the conditions in the research pattern.

When the results of the main study problem were examined, the findings that were to be considered for performing the CDM analysis were obtained. The effect of missing data on the test parameters was observed to be variable when the density of the itemóattribute relation was considered. The stochastic element of the DINA model ensures that the response behavior is not deterministic but probabilistic; this occurs because the respondents make false-positive and false-negative errors at an item level while responding. Therefore, the possibility of accurately responding to an item is determined by two different error probabilities (Rupp, & Templin, 2008). Thus, the g and s parameters in the CDM are associated with false-positive and true-negative errors in various models. The most common definition of the g parameter corresponds to the probability of individuals who do not have at least one of the attributes required to accurately answer an item. Further, the s parameter corresponds to the false response possibilities of the individuals who possess all the required attributes (de La Torre, & Douglas, 2004). These two possibilities have been separately discussed in this research. First, the observation of missing data on the items related to a single attribute (B1) more adversely affected the g parameter estimation when compared with the missing data located in items related with more attributes (B2 and B3). In other words, as the number of attributes related to an item increased, the missing data in these items minorly affected the consistency level of the g parameter of the test. However, the situation was observed to be the opposite for the s parameter. The highest consistency in the s parameters was observed when the missing data were located in the items associated with a single attribute. Further, the missing data in the items associated with a single attribute were observed to minorly affect the s parameter consistency level. This can be mainly attributed to the false-positive and true-negative property of the g parameter, as described above. Each true-negative decision that occurred during the missing data imputations for items that were associated with large number of attributes changed the g parameter for each attribute. However, the true-negative decision in this item was observed to affect only that item. This situation was observed to occur in the opposite manner for the s parameter. The s parameter was also basically determined using the false-negative and true-positive decisions. The s parameter exhibits the possibility of a non-mastery decision about an examinee who essentially possesses the necessary attributes for answering an item. Further, the false-negative possibility increases as the itemóattribute relation increases.

One of the remarkable findings observations of this research was the amount of change in latent class assignments in the presence of missing data. The latent class consistency decreased as the amount of missing data increased. However, as the number of attributes associated with an item increased, the latent class consistency was observed to be less affected by the missing data. In other words, the latent class consistency, which was observed in items exhibiting high itemóattribute relations, was higher when compared with that in the items with less itemóattribute relation in the presence of missing data. Thus, it can be said that the highly saturated items may have a minor negative impact on the missing data studies because of the effect of the g parameter while determining the latent classes.

Conducting the research using a single CDM model results in a limitation with respect to the generalizability of the findings. However, some information has been obtained in case of

CDM studies while considering the consistency of findings and the significance of the differences. The research results denoted that the test developers who use the CDM models should consider the relation between an item and the attribute in case of items with missing data. In future studies, verifying the conditions of variables, such as item discrimination, item difficulty, test length, sample size, ability level of the sample, and the application of different CDM models, will contribute to the literature to reveal the effects of the missing data and the itemóattribute relation better.

References

Ba okçu, T. O. , Kalkan, Ö. K., & Ö retmen, T. (2016, September). *DINA Modele Dayal, Madde Parametre Kestiriminde Kay,p Veri Ele Alma Yöntemleri Etkisinin ncelenmesi.* [An Investigation of the Effect of Missing Data Handling Methods on Parameter Estimation Based on DINA Model]. Paper presented at the Fifth International Congress on Measurement and Evaluation in Education and Psychology, Antalya, Turkey. p134-135. Abstract retrieved from http://epod2016.akdeniz.edu.tr/_dinamik/333/53.pdf

Bennett, D. A. (2001). How can i deal with missing data in my study? *Australian and New Zealand journal of public health*, *25*(5), 464-469.

Buuren, S. V., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of statistical software*, 45(3), 1-67.

Chen, J. (2017). A Residual-Based Approach to Validate Q-Matrix Specifications. *Applied Psychological Measurement*, (135), 014662161668602. https://doi.org/10.1177/0146621616686021

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. (2011). The generalized DINA model framework. *Psychometrika*, *76*(2), 179-199.

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., & Douglas, J. A. (2008). Model evaluation and multiple strategies in cognitive diagnosis: An analysis of fraction subtraction data.*Psychometrika*, *73*(4), 595-624.

de La Torre, J., Hong, Y., & Deng, W. (2010). Factors affecting the item parameter estimation and classification accuracy of the DINA model. *Journal of Educational Measurement*, *47*(2), 227-249.

de la Torre, J., & Lee, Y. S. (2010). A Note on the Invariance of the DINA Model Parameters. *Journal of Educational Measurement*, *47*(1), 1156127. https://doi.org/10.1111/j.1745-3984.2009.00102.x

Dong, Y., & Peng, C. Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus*, *2*(1), 222.

Doornik, J. A. (2018). An Object-Oriented Matrix Programming Language Ox 8. Timberlake Consultants.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah. NJ: Erlbaum.

Enders, C. K. (2010). Applied missing data analysis. Guilford Press.

Finch, H. (2008). Estimation of item response theory parameters in the presence of missing data. *Journal of Educational Measurement*, *45*(3), 225-245.

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica*, *37*(6), 3596374. https://doi.org/10.1016/0001-6918(73)90003-6.

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological methods*, *11*(4), 323.

Haertel, E. H. (1989). Using Restricted Latent Class Models to Map the Skill of Achievement Structure Items. Journal *of Educational Measurement*, *26*, 3336352. http://dx.doi.org/10.1111/j.1745-3984.1989.tb00336.x

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

Henson, R., & Douglas, J. (2005). Test construction for cognitive diagnosis. *Applied Psychological Measurement*, *29*(4), 262-277.

Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, *74*(2), 191.

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, *25*(3), 258-272.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data.* Wiley. New York.

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika, 64*(2), 187-212.

Nichols, P. D., Chipman, S. F., & Brennan, R. L. (2012). *Cognitively diagnostic assessment.* Routledge.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of educational research*, *74*(4), 525-556.

R Core Team. (2018). R: A language and environment for statistical computing [Computer Software]. Vienna, Austria: R Foundation for Statistical Computing.

Robitzsch, A., Kiefer, T., George, A. C., Uenlue, A., & Robitzsch, M. A. (2018). Package ÷CDMø

Rupp, A. A., & Mislevy, R. J. (2007). Cognitive Foundations of Structured Item Response Models. In Leighton, J., Gierl, M. (Eds.). *Cognitive diagnostic assessment for education: Theory and applications, 205-241*. New York: Cambridge University Press.

Rupp, A. A., & Templin, J. (2008). The effects of Q-matrix misspecification on parameter estimates and classification accuracy in the DINA model. *Educational and Psychological Measurement*, *68*(1), 78-96.

Schlomer, G. L., Bauman, S., & Card, N. A. (2010). Best practices for missing data management in counseling psychology. *Journal of Counseling psychology*, *57*(1), 1.

Sen, S., & Bradshaw, L. (2017). Comparison of Relative Fit Indices for Diagnostic Model Selection. *Applied Psychological Measurement*, 014662161769552. https://doi.org/10.1177/0146621617695521

Sijtsma, K., & Van der Ark, L. A. (2003). Investigation and treatment of missing item scores in test and questionnaire data. *Multivariate Behavioral Research*, *38*(4), 505-528.

Sorrel, M. A., Olea, J., Abad, F. J., de la Torre, J., Aguado, D., & Lievens, F. (2016). Validity and Reliability of Situational Judgement Test Scores. *Organizational Research Methods*, *19*(3), 506ó 532. https://doi.org/10.1177/1094428116630065

Stekhoven, D. J. (2016). MissForest: nonparametric missing value imputation using random forest. *R package version 1.4*.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Allyn & Bacon/Pearson Education.

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://dx.doi.org/10.1111/j.1745-3984.1983.tb00212.x

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological methods*, *11*(3), 287.

van der Linden, W. J., & Hambleton, R. K. (2013). Item Response Theory: Brief History, Common Models, and Extensions. In *Handbook of Modern Item Response Theory*. https://doi.org/10.1007/978-1-4757-2691-6_1

von Davier, M. (2005). *A general diagnostic model applied to language testing data* (ETS Research Report RR-05-16). Princeton, NJ: Educational Testing Service.

Winship, C., Mare, R. D., & Warren, J. R. (2002). Latent class models for contingency tables with missing data. *Applied latent class analysis*, 408.

Zhang, B., & Walker, C. M. (2008). Impact of missing data on personô model fit and person trait estimation. *Applied Psychological Measurement*, *32*(6), 466-479.

Geni letilmi Özet

Giri

E itimsel ve psikolojik ölçmelerde, Bili sel Tan, Modelleri'ne (BTM) olan ilgi son y,llarda dikkate de er bir art, göstermektedir. Klasik Test Kuram, (KTK) ve Madde Tepki Kuram, (MTK) kat,l,mc,lar,n genel yeteneklerini de erlendirmeye çal, ,rken, BTM'ler, belirli bir özellik kümesine dayanarak adaylar,n zay,f ve güçlü yönlerini belirlemeye odaklan,r. Bu nedenle, CDM'ler s,navlarda adaylar,n tek bir genel s,nav puan,n, rapor etmek yerine, belirli becerilerdeki uygulamalar,n, gösteren daha ayr,nt,l, bir de erlendirme sunarlar (Sen ve Bradshaw, 2017). Literatürde birçok CDM modeli bulunmas,na ra men, basit formülü, yorumlama kolayl, , ve iyi model-veri uyumu sa lamas, nedeniyle DINA model birçok ara t,rmac, taraf,ndan tercih edilmektedir (de la Torre ve Douglas, 2008; de la Torre ve & Lee, 2010, Rupp ve Templin, 2008). DINA modeli s,n,fland,rma do rulu u ve madde parametresi tahminleri; Q matris yap,s,, örtük s,n,f yap,s,, önsel karakteristi i, örneklem büyüklü ü, tahmin ve kayd,rma parametre de erleri ve kestirim yöntemi gibi çe itli faktörlerden etkilenmektedir (De La Torre, Hong, & Deng, 2010). Bunlara ek olarak, kay,p veriler parametre tahminlerinin do rulu unu ve tutarl,l, ,n, etkileyebilir ve örtük s,n,f modellerinde yanl, kestirimlere yol açabilirler (Winship, Mare ve Warren, 2002). E itimsel ve psikolojik testlerle bireylerden veri topland, ,nda, kay,p verilerle kar ,la ,lmas, muhtemeldir. Kay,p veriler di er birçok faktör gibi, DINA model madde ve örtük s,n,f kestirimlerini etkileme potansiyeline sahiptir (Ba okçu, Kalkan ve Ö retmen, 2016). Bu nedenle, eksik veri varl, ,nda madde parametresi ve örtük s,n,f kestirimlerinin tutarl,l, ,n, incelemek önemli görünmektedir. Ayr,ca, madde-özellik ili kisinin kay,p verilerden nas,l etkilendi inin belirlenmesi, bili sel tan,ya dayal, testlerin olu turulmas,na önemli katk,lar sa layacakt,r. Bu ara t,rman,n amac,; farkl, miktarlarda kay,p veri varl, ,nda madde-özellik ili kisinin, DINA model madde parametre ve örtük s,n,f kestirimlerini nas,l etkiledi ini incelemektir.

Yöntem

Ara t,rma kapsam,nda özellik-madde ili kileri simülasyon verileri ile incelenmi tir. Bu amaçla; veri üretimi ve parametre kestirimleri için õCDMö (Robitzsch, Kiefer, Uenlue, & Robitzsch, 2018), veri silme için õmissForestö (Stekhoven, 2016), Multiple Imputation (MI) için õmiceö (van Buuren, & Groothuis-Oudshoorn, 2011), R (R Core Team, 2018) paketleri kullan,lm, t,r. Verilerin üretilmesinde 4 özellik ve 24 maddeden olu an bir Q matris kullan,lm, t,r. Q matris yap,land,r,l,rken ilk 8 madde (1-8, [block1, B1]) 1, sonraki 8 madde (9-16, [block2, B2]) 2, son 8 madde (17-24, [block3, B3]) ise 3 özellik ile ili kilendirilmi tir. Böylece her bir blokta (B1, B2, B3) ve toplamda her bir özellik ile ili kili e it say,da maddeye sahip bir Q matris elde edilmi tir. Ba lang,çta g ve $s$nin 0.1-0.3 aras,nda de i ti i 3000 ki ilik bir veri seti üretilmi tir. Bu veri setleri analiz edilerek guess (g) ve slip (s) parametreleri ve örtük s,n,f kestirimleri elde edilmi tir. Elde edilen bu parametreler gerçek de er olarak kabul edilmi tir. Ara t,rma kapsam,nda %5, %10 ve %15 olmak üzere 3 farkl, kay,p miktar, belirlenmi tir. Ba lang,çta üretilen veri setinden ilk 8 (B1), orta 8 (B2) ve son 8 (B3) madde için s,ras, ile rassal %5, %10 ve %15 veri silinmi tir. Ard,ndan, bu veri setlerine MI yöntemi ile imputasyon (n=5) gerçekle tirilmi tir. Bu i lemler, her bir ko ul için 100 kez tekrarlanm, t,r. Daha sonra, bu veri setlerinden elde edilen g ve s parametreleri ve örtük s,n,f kestirimleri kay,ps,z veri setinden elde

edilen gerçek de erler ile kar ,la t,r,lm, t,r. Bu kar ,la t,rmalar ortalama mutlak fark ve uyum oranlar, üzerinden yap,lm, t,r.

Bulgular ve Tart, ma

Bu çal, ma farkl, oranlarda kay,p veri varl, ,nda, madde-özellik ili kisinin, DINA model madde parametre ve örtük özellik kestirimlerini nas,l etkiledi ini ele almaktad,r. Kay,p veri oran, 5% iken B1, B2 ve B3 bloklar,ndan elde edilen *g* parametre RMSE de erleri .0004-.005 aral, ,nda, *s* parametre RMSE de erleri .001-.0085 aral, ,nda de i mi tir. Örtük s,n,f kestirim uyum ortalamalar, ise .95-.98 aral, ,nda de i mi tir. Bu bulgu Tabachnick ve Fidell (2007) taraf,ndan rapor edilen büyük örneklemlerde 5% veya daha az kay,p verinin önemli sorunlara yol açmayaca , bulgusunu desteklemektedir. Kay,p veri miktar,n,n artmas, *g-s* parametre RMSE de erlerinde bir art, a neden olurken, örtük s,n,f kestirimleri uyum de erlerinde ise bir dü ü e neden olmaktad,r. Kay,p veri art, , ile ilgili bu sonuç asl,nda beklenen bir durumdur. Kay,p veri miktar,n,n artmas, bütün modeller için ölçme sonuçlar,n,n tutarl,l, ,n, olumsuz yönde etkileyen bir faktördür. Bu durum hem madde parametrelerindeki de i imde hem de örtük s,n,flar,n uyum düzeyinde gözlenmektedir.

Ara t,rmadan elde edilen bir di er bulgu ise kay,p verinin gözlendi i maddeler ile testin genel parametrelerinin etkilenme düzeylerinin ili kili olmas,d,r. Analizler sonucunda beklenildi i gibi kay,p verinin bulundu u maddelere ili kin parametreler, testte yer alan di er maddelerin parametrelerine göre önemli düzeyde gerçek de erlerinden uzakla maktad,r. Elde edilen bulgular incelendi inde ara t,rma deseninde yer alan bütün ko ullar için kay,p verinin yer ald, , blo a ait parametre de i iminin di er bloklardan daha yüksek oldu u görülmü tür.

Ara t,rman,n as,l problemine ili kin sonuçlar incelendi inde, CDM analizleri için dikkate al,nmas, gereken bulgulara ula ,ld, , görülmektedir. Madde özellik ili kisinin yo unlu u göz önüne al,nd, ,nda kay,p verinin test parametrelerine etkisi de de i kenlik göstermi tir. Öncelikle *g* parametreleri incelendi inde tek özellikle ili kili maddelerde (B1) kay,p verinin gözlenmesinin, daha çok özellikle ili kili maddelerde yer alan kay,p veriye göre (B2, B3) parametre kestirimini daha fazla olumsuz etkiledi i görülmü tür. Bu durumun temel nedeni *g* parametresinin false pozitif ve true negatif özelli inden kaynaklamaktad,r. Daha fazla özellik ile ili kili olan maddelere yönelik yap,lan kay,p veri atamalar,nda gerçekle en her true negatif karar her bir özellik için *g* parametresinde de i ime yol açmaktad,r. Ancak, bu durum *s* parametresi için tam tersi bir da ,l,m göstermektedir. *s* parametrelerinde en yüksek tutarl,l,k bir özellik ile ili kili olan maddelerde kay,p veri gözlendi inde ortaya ç,kmaktad,r. *s* parametresi de temelde false negatif ve true pozitif kararlar, ile belirlenmektedir. *s* parametresi madde için temelde gerekli özelliklere sahip olan cevaplay,c, hakk,nda õzelliklere sahip olmad, ,ö karar, verilmesi durumlar,n,n bir olas,l, ,n, göstermektedir. Bu noktada madde özellik ili kisi artt,kça false negatif olas,l, , da artmaktad,r.

Ara t,rman,n dikkat çekici bulgular,ndan biride kay,p veri varl, ,nda örtük s,n,f atamalar,ndaki de i im miktar,d,r. Örtük s,n,f uyumlar, kay,p veri miktar, artt,kça dü mektedir. Ancak maddenin ili kili oldu u özellik say,s, artt,kça örtük s,n,f uyumu kay,p veriden daha az etkilenmektedir. Bu durum özellikle *g* parametresinin örtük s,n,flar, belirlemekteki etkisi ile birlikte daha doygun maddelerin kay,p veri çal, malar,nda sonuçlar, daha az düzeyde olumsuz yönde etkileyece ine i aret etti i de söylenebilir.

Ara t,rman,n tek bir CDM modeli üzerinde yürütülmesi ula ,lan bulgular,n genellenebilirli i aç,s,ndan bir s,n,rl,l,k te kil etmekle birlikte, bulgular,n kararl,l, , ve farklar,n anlaml,l, , göz önüne al,nd, ,nda, CDM çal, malar, için üzerinde durulmas, gereken baz, bilgilere ula ,ld, , da iddia edilebilir. Ara t,rman,n sonuçlar, özellikle CDM modellerini kullanan test geli tiricilerinin kay,p veri gözlenen maddelerde, madde-özellik ili kisini göz önünde bulundurmalar, gerekti ini ortaya koymaktad,r.