ISSN: 1309 – 6575
Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi
Journal of Measurement and Evaluation in Education and Psychology
Research Article; 2023; 14(4); 413-439

A Comparison of the Classification Performances of the DINO Model, Artificial Neural Networks and Non-Parametric Cognitive Diagnosis

Emine YAVUZ* Hakan Yavuz ATAR**

Abstract

The purpose of this study was to compare the attribute (ACR) and pattern-level (PCR) classification rates of the Deterministic-Input, Noisy-Or Gate (DINO) model, Artificial Neural Networks (ANNs), and Non-Parametric Cognitive Diagnosis (NPCD) on simulation datasets. As a comparison condition, the number of attributes, sample size, the number of items, and missing data rate were chosen. A further purpose was to examine the similarities between the classification rates of the DINO model, ANNs, and NPCD on the PISA 2015 collaborative problem-solving (CPS) datasets in various numbers of attributes and sample sizes. For the study, simulation datasets were generated on the basis of the complex Q matrix structures and the DINO model. The conditions for the sample size factor for the real datasets were determined by simple random selection among the participants in the PISA 2015 administration. As a result, it was found that there was a similarity between the DINO model and NPCD classification rates in both simulation and real datasets. In addition, regarding the increase in sample size in both simulation and real datasets, no consistency was found in the increase or decrease of the classification rates of ANNs and NPCD and the similarities of these rates.

Keywords: diagnostic assessment, non-parametric cognitive diagnosis, artificial neural network, DINO, PISA

Introduction

In recent years, the importance of assessment and evaluation of twenty-first-century skills has increased. In parallel with this, the "No child left behind" (2001) act has caused a shift in assessment methods from summative to more diagnostic and formative. In this context, cognitive diagnosis models have started to be used in modeling and evaluating the complex structures of twenty-first-century skills and provide students with detailed feedback formed within the framework of diagnostic and formative assessment. The placement of students into attribute classes in cognitive diagnostics is generally performed through parametric Cognitive Diagnosis Models (CDMs).

When CDMs were first being developed, CDMs estimation algorithms were not publicly available (Chiu et al., 2017), and they usually required large samples as well as involving complex computational calculation procedures that could only be carried out through relatively expensive software, which restricted the use of CDM applications extensively (Chiu & Douglas, 2013; Chiu & Köhn, 2019). In addition, CDM analysis results were found to yield biased results in some situations with complex structures (i.e., a high number of attributes) or fewer items (Shu et al., 2013; Shuying, 2016). Hence, researchers have embarked on a quest to find a non-parametric method to be used in cognitive diagnosis. Furthermore, the need for the development and use of non-parametric cognitive diagnostic techniques has recently increased because these models were found to provide promising results with smaller

*Asst. Prof. Dr.,Erciyes University, Faculty of Education, Kayseri-Türkiye, emineyavuz@erciyes.edu.tr, ORCID: 0000-0002-1991-1416
** Prof. Dr., Gazi University, Faculty of Education, Ankara-Türkiye, hakanyavuzatar@gmail.com, ORCID ID: 0000-0001-5372-1926

To cite this article:
Yavuz, E. & Atar, H.Y. (2023). A Comparison of the classification performances of the DINO model, artificial neural networks and non-parametric cognitive diagnosis. *Journal of Measurement and Evaluation in Education and Psychology, 14*(4), 413-439. https://doi.org/10.21031/epod.1246752

Received: 2.02.2023
Accepted: 20.10.2023

samples and a high number of attributes (Paulsen & Valdivia, 2022). Therefore, this study has focused on the non-parametric perspective in cognitive diagnosis.

Cognitive diagnosis involves non-parametric methods such as the Attribute Hierarchy Method (AHM), clustering techniques (such as k-means and hierarchical agglomeration), Artificial Neural Networks (ANNs), and Non-Parametric Cognitive Diagnosis (NPCD). Since there was no hierarchy among the structure attributes measured in the real dataset, the AHM was not used in this study. Furthermore, clustering techniques that have an exploratory perspective were not included, as the study was carried out in the context of cognitive diagnosis with a confirmatory perspective. In this regard, the present study focused on the NPCD and ANN models used in cognitive diagnostic classification. These methods are similar to CDMs as they require knowledge about the data structure (compensatory or non-compensatory) and Q matrix for the analysis, but they differ from each other in that they make use of different perspectives to classify students according to their attribute profiles.

Non- Parametric Cognitive Diagnosis (NPCD)

Students are categorized in NPCD by comparing their observed response vectors to the ideal response patterns (Chiu & Douglas, 2013). Observed response vectors are the vectors of students' dichotomous responses to the test, and the observed response of a student i is usually represented by y_i. Ideal response patterns are the item response patterns that are expected to occur as a result of the comparison between the attributes that students master and the attributes required for students to respond to the items correctly. It is necessary to know the compensatory state (compensatory or non-compensatory) of the data structure for estimating the ideal response patterns. For the DINO model, which is one of the compensatory models, the ideal response of the student i to item j is calculated by using equation 1:

$$\omega_{ij} = 1-\prod_{k=1}^{K}(1-\alpha_{ik})^{q_{jk}} \qquad (1)$$

In equation 1, α_{ik} is whether the student i has attribute k or not, and q_{jk} is whether the presence of item j requires attribute k. A test with k^{th} attribute involves m=1,..,2^k attribute profiles; therefore, all ideal responses possible for α_i can be stated as $\omega_m=\omega_1, \ldots, \omega_{2^k}$. Since ω_i is determined by α_i, the distance between the observed (y_i) and ideal response patterns of the student i with attribute α_m profile can be represented as d(y_i, α_m). The estimated attribute pattern ($\hat{\alpha}_i$) in NPCD is the attribute pattern that minimizes the distance between all ideal item response patterns and the observed response patterns of student i (Chiu & Douglas, 2013). In other words, $\hat{\alpha}_i$ can be defined as the attribute pattern of the ideal response pattern, which is the most proximate or similar to the observed item response pattern among all ideal item response patterns (Chiu et al., 2017).

$$\hat{\alpha}_i = \arg\min D(y_i, \alpha_m) \quad m \in (1,2,\ldots,2^k) \qquad (2)$$

NPCDs estimate the attribute classes by comparing the observed item response patterns with each of the ideal response profiles of the possible 2^k attribute classes (Chiu et al., 2017). Various distance measures (Hamming, weighted Hamming, and penalized Hamming) can be used to measure the similarity between two vectors. (Chiu & Douglas, 2013). It has been determined that the weighted Hamming distance is more efficient, reduces the number of links between possible ideal response patterns, and yields higher classification accuracy results (Paulsen, 2019). Therefore, weighted Hamming was used as the distance measure in this study. The Hamming distance and the weighted Hamming distance for the compensatory data are calculated with equations 3 and 4, respectively:

$$d_h(y_i, \alpha_m) = \sum_{j=1}^{J} |y_{ij} - \omega_{mj}| \tag{3}$$

$$d_{\mathrm{wh}}(y_i, \alpha_m) = \sum_{j=1}^{J} \frac{1}{\overline{p}_j(1-\overline{p}_j)} |y_{ij} - \omega_{mj}| \tag{4}$$

Multilayer Perceptron Artificial Neural Networks (MLP-ANN)

These are the models inspired by the structures and functions of neurons in the human brain. After Gierl et al. proposed to use ANNs together with AHM in 2007, ANNs started to be used within the scope of CDM. ANNs are employed in the context of CDM through two different learning paradigms: supervised and unsupervised learning paradigms. We utilized the supervised learning paradigm since it is compatible with the confirmatory perspective of CDMs.

The multilayer perceptron ANN (MLP-ANN) can be defined as a parallel processing architecture (Garson, 1998; Gierl et al., 2007, 2008) that receives stimuli with input units and carries these to the output unit with latent units. An input layer, at least one latent layer, and an output layer are typically present in these ANNs. A variety of neurons with various roles make up each layer. The latent layer(s) in the network helps to model the effects of the interaction of input neurons on output neurons. In the MLP-ANN, neurons in a particular layer cannot interact with each other directly, but they can only connect with neurons in the adjacent layers. These connections are called weights. The process of estimating the weights in an ANN or correlating the input layer with the output layer is called 'ANN training' or 'ANN learning'. The training of the network is carried out iteratively in such a way that each iteration trains the network to minimize the difference (error) between the expected and observed attribute values for an estimated response pattern. In this iterative method, weight estimations are initialized with arbitrary values drawn from the ordinary normal distribution (the default approach of the R package neuralnet). In this study, weight back-tracking was used for the weight estimations in the framework of the resilient back-propagation approach. After the training of the ANN is completed, the cognitive diagnosis is terminated by analyzing a new dataset in which the inputs are known, but the outputs are not known.

Within the scope of CDMs, each input layer neuron of ANNs represents an item that constitutes the test, and the responses given to those items are used as input. The neurons of the output layer are interpreted as the attributes that the predetermined test is intended to measure. The expected response patterns derived from the Q matrix, or the collection of ideal response vectors produced from the Q matrix, are the inputs utilized to train the ANN. The results are the pertinent attribute profiles based on the validity of the Q matrix (Cui et al., 2016; 2017). Training of the network refers to the establishment of the connection between these ideal response vectors and the relevant attribute profiles. The training process continues until the neural network learns the connections. After the network has completed the learning process, students' responses to the test items are put into the network as input for analysis. It is possible to summarize the mathematical form of an ANN training with the following steps. If it is assumed that there is a three-layer MLP-ANN where i refers to input neurons, j to latent neurons, and k to output neurons, the weighted sum of all input neurons is obtained in the first step.

$$a_j = \sum_{j=1}^{J} W_{ji} X_i \qquad (5)$$

a_j, is the weighted sum for latent neuron j, W_{ji} is the weight of the connection from input neuron i to latent neuron j, and X_i is the value of input neuron i. This sum is converted to the following form by the activation function f(·) to calculate the latent neuron value:

$$h_j = \mathrm{f}(a_i) = f\left(\sum_{j=1}^{I} W_{\mathrm{ji}} X_i\right) \qquad (6)$$

Since the logistic/sigmoid function is frequently preferred as the activation function in the context of cognitive diagnosis (Guo et al., 2017), the logistic function was used in this study in the same way to determine the weight of two connections. When the latent neuron values have been obtained, the output neuron values must be calculated. The output neurons are calculated in a way that is similar to the calculation of the latent neurons.

The association of the input neurons with the output neurons through two successive conversions continues until the error function is reduced to the minimum or to a predetermined value or until it is fixed. In other words, the values of the connection weights that minimize the error function are selected for the estimation of the model parameters. In this study, the cross-entropy value, which is an error calculation function, was examined, and the number of hidden layers and neurons was determined. Since only MLP-ANN was studied within the scope of this research, ANN refers to MLP-ANN in this study.

There are some studies in which classification rates of ANNs and NPCDs are examined together under various conditions (see McCoy & Willse, 2014; Paulsen, 2019; Paulsen & Valdivia, 2022) in the related literature. The current research has some similarities and differences with the study of McCoy and Willse (2014) and Paulsen (2019). The similarities include the binary coding of attributes (0-1) and the multivariate normal distribution of attributes in simulation data. Since the data in the study of McCoy and Willse (2014) and Paulsen (2019) were created based on the Deterministic-Input, Noisy-And Gate (DINA) model, the findings obtained from the studies can be generalized to non-compensatory data structures. The use of both the real and DINO-based simulation datasets and the analysis of the classification performances of compensatory models based on the data structures are the features that distinguish the present study from the previous studies. The current study also differed in the number of attributes used in data analysis. While McCoy and Willse's (2014) and Paulsen's (2019) studies had a maximum of eight attributes, the current study examined a maximum of 11 attributes. Additionally, in the studies of McCoy and Willse (2014) and Paulsen (2019), item discrimination was evaluated as high or low, whereas in the current study, item discrimination was handled at a moderate level, based on the recommendation of the previous studies (see Guo et al., 2017; Shuying, 2016). Finally, the current research differs from the previous studies in that it examines the missing data effect in the field of non-parametric cognitive diagnostics and the effect of NPCD and ANN on classification rates. It is expected that comparing the findings of McCoy and Willse (2014) and Paulsen (2019) with the findings obtained from the simulation datasets generated based on real and compensatory models in this study will provide a holistic perspective regarding the effects of various factors on classification in non-parametric cognitive diagnosis.

Finally, this study is considered important as it provides a basis for further studies to be conducted in this field, especially on the use of NPCD and ANNs in evaluating students' twenty-first-century skills and comparing the classification performances of these methods. In light of this, the aims of this research are to compare the attribute and pattern-level classification rates of the DINO models, ANN, and NPCD in various settings in DINO-based simulation datasets and then to look at the similarities of the classification rates of the DINO models, ANN, and NPCD in the PISA 2015 CPS dataset. For this purpose, the study sought answers to the following questions:

- Do the attribute and pattern-level classification rates of the DINO models, ANN, and NPCD in simulation data differ according to the number of attributes (3, 5, and 7), sample sizes (30, 100, and 500), the number of items (15, 30, and 45) and the missing data rates (0, .05, and .10)?
- What is the similarity of the classification rates of the DINO models, ANN, and NPCD in the real data (PISA 2015 CPS) according to the number of attributes (3, 7, and 11) and sample sizes (30, 100, and 500)?

Methods

Research Design

Since this research aims to determine the classification performances of the DINO model, ANN, and NPCD under different settings, this study was designed as a simulation study. Simulation studies are frequently used to assess the performance of a specific statistical model or to predict the results of a given situation.

The Population

The PISA 2015 CPS administration data was utilized in the study due to the convenience of obtaining the Q matrix. A total of 414,498 students from 52 countries participated in the CPS assessment. Since Form 93 and Form 96 in the PISA 2015 CPS administration included more items representing the CPS structure, Form 93 and Form 96 were used in the study. Furthermore, the items scored as polytomous previously were rescored as dichotomous to maintain the content validity. After removing the missing data, 18,170 students from 43 countries were used in the data analysis. Using a simple random sampling procedure in IBM SPSS 26.0, sample sizes of 30, 200, and 500 students were sampled from the population of 18,170 students.

Research Procedure

The conditions of the simulation and the real datasets are explained in detail in the following sections.

Obtaining Simulation Datasets

In order to ensure the compatibility of the simulation datasets with the structure of the real datasets, simulation datasets were generated based on the DINO model, which is a frequently used compensatory model. The first step of the data generation process was the determination of the Q matrix structures. In the creation of the Q matrices, it was ensured that each attribute was measured by the same number of items and that the items measuring more than one attribute were also equal in number (see de la Torre, 2008; de la Torre & Douglas, 2008; Rupp & Templin, 2008a). An item was allowed to measure a maximum of three attributes in each Q matrix. In addition, the increase in the complexity of all the Q matrix structures used in the study indicated an increase in the number of attributes. Finally, all the Q matrices were completed (Chiu et al., 2009). Within the scope of the research, firstly, a 15-item Q matrix was created, then this structure was used twice to create the 30-item Q matrix and three times to create the 45-item Q matrix. The 15-item Q matrices of the simulation datasets are reported in Table 11 in the appendices.

After the determination of the Q matrix structures, features of the attributes were identified. Multivariate normal distribution was used in generating the attributes of the simulation data for describing the realistic situations in which the attribute patterns were not equally distributed and the attributes were correlated. The correlation value between the attributes was determined to be $\rho = .5$ (see Chiu & Douglas, 2013; McCoy & Willse, 2014).

After the distributions of attributes and the level of correlation between the attributes were determined, the item parameters (s and g parameters) were identified. It was found that the s and g parameter values were between .33 and .35 when the number of attributes in the real datasets used in the study (PISA 2015 CPS competency) was 3, 7, and 11. In order to make the structure of the simulation datasets as similar as possible to the structure of the real datasets and not to exceed the reliability of classification accuracy, s and g parameter values in all conditions were determined as U [0, .4] in uniform distribution.

The factors addressed in the study were the number of attributes (3, 5, and 7), sample sizes (30, 100, and 500), number of items (15, 30, and 45), and missing data rate (0, 5%, and 10%). These factors and their conditions were selected based on the literature review due to their frequent use in the simulation and real data research and the rich information they provide to the implementers and the readers. Within the scope of the study, the comparison condition of 3x3x3x3x3=243 was created in the simulation datasets, and 100 replications were carried out for each comparison condition of the simulation data.

Data collection tools for real datasets and data collection

In this study, the PISA 2015 CPS data were used due to the accessibility of the information about the Q matrix as the real datasets. The PISA 2015 CPS competency is a structure that is formed by the combination of collaboration skills and PISA 2012 individual problem-solving process skills. The skills constituting the PISA 2015 CPS competency and details about the skill(s) measured with these items were described in OECD reports (2017a; 2017b). Q matrices were primarily created by making use of these reports based on expert opinions, and the real datasets were analyzed. The Q matrices of the first 17 items are reported in Table 12 in the appendices. After Q matrix validation was performed with the PVAF method (de la Torre & Chiu, 2015), the analyses of the real datasets were performed again, and the findings obtained from both types of the Q matrices were interpreted together. The factors used in the analysis of the real datasets were the number of attributes (3, 7, and 11) and the sample sizes (30, 100, and 500). Furthermore, the comparison condition of 3x3x3=27 was created in the real datasets.

Data Analysis

We used R-Studio for the data generation and the validation and analyses of the Q matrices of the real datasets, and IBM SPSS 26.0 software (IBM, 2019) for the factorial ANOVA (analysis of variance) in the study. Moreover, various packages were used for different purposes: GDINA 2.8.0 (Ma et al., 2020) was used for the validation of the Q matrices of the real data; the packages of CDM 7.4-19 (Robitzsch et al., 2019) for DINO analysis; the packages of NPCD 1.0-11 (Zheng et al., 2019) for the NPCD analysis; the packages of neuralnet 1.44.2 (Fritsch et al., 2019) for ANN analysis; the packages of missForest 1.4 (Stekhoven, 2016) for the missing data creation, and TestDataImputation 1.1 (Dai et al., 2019) for missing data imputation.

The simulation datasets were analyzed by following the procedures of NPCD, DINO model, and ANN analyses to answer the first research question, while the classification performances were determined in various conditions. For the simulation dataset, the classification performances of the NPCD, DINO model, and ANN were investigated through ACR and PCR. The equations of ACR and PCR are presented below.

$$ACR = \frac{1}{NxK} \sum_{i=1}^{N} \sum_{i=1}^{K} I[\widehat{\alpha_{ik}} = \alpha_{ik}] \quad (7)$$

$$PCR = \frac{1}{N} \sum_{i=1}^{N} I[\widehat{\alpha_{i}} = \alpha_{i}] \quad (8)$$

where N is the sample, K is the number of attributes, i is a student, α_{ik} is the k. attribute that the student i. actually has, and $\hat{\alpha}_{ik}$ is the k. attribute that the student i. is estimated to have.

After the ACRs and PCRs were obtained for the DINO model, ANN, and NPCD, factorial ANOVAs were carried out, and the effect sizes were interpreted according to Cohen (1988). Cohen (1988) classified effect sizes as small (.20), medium (.50), and large (.80).

The analyses of the same real datasets (PSA 2015 CPS) were performed by making use of two different Q matrices that were created based on the technical reports and the PVAF method to address the second research question. ACRs and PCRs of the DINO model ANN and NPCD could not be calculated due to the fact that the real attribute classes of the students were not known in the real datasets. As a result, equations 11 and 12 were modified in accordance with the real data analyses, and the similarity of the attribute (SACR) and pattern-level classification rates (SPCR) of the DINO model, ANN, and NPCD were obtained. In this regard, α_{ik} was replaced with α value estimated by one of the compared models and $\hat{\alpha}_{ik}$ with α value estimated by the other model.

Results

1.Comparison of the ACRs and PCRs of the DINO model, ANN, and NPCD in the simulation data based on the number of attributes (3, 5, and 7), sample sizes (30, 100, and 500), number of items (15, 30, and 45) and missing data rate (0, .05, and .10)

The ACRs and PCRs of the DINO model, ANN, and NPCD under study conditions are given in Table 13 in the appendices.

1.1. Comparison of the ACRs and PCRs of the DINO modes, ANN, and NPCD according to the number of attributes (3, 5, and 7)

The results of the ACRs of the DINO model, ANN, and NPCD, and the comparison of the number of attributes (3, 5, and 7) factors are presented in Table 1. It can be seen in Table 1 that the interaction between the number of attributes and the models showed a statistically significant difference ($p<.01$) and that this interaction had a medium effect (η_p^2 = .39).

Table 1

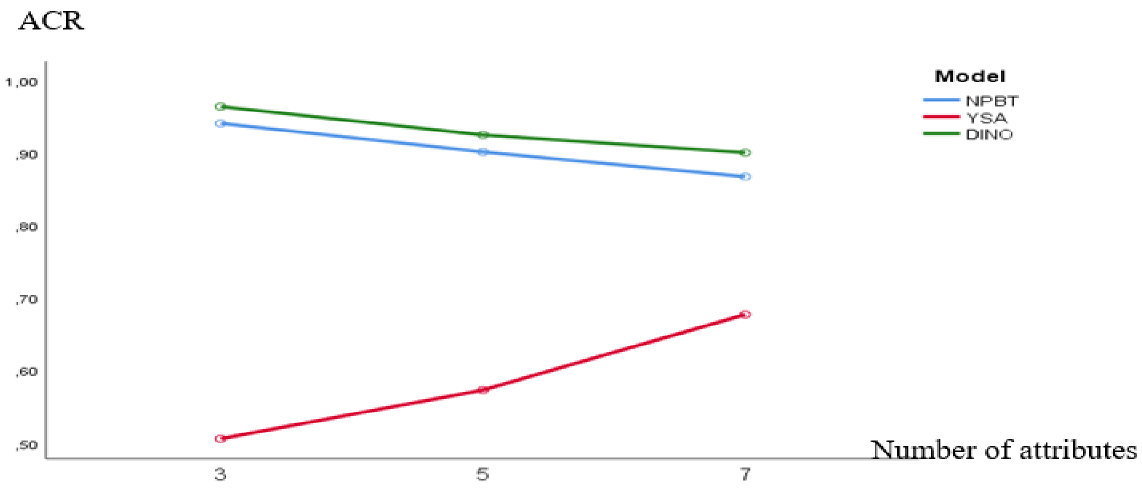
Factorial ANOVA Results of the DINO Model, ANN, and NPCD's ACRs according to the Number of Attributes

Source of variation	Sum of squares	*df*	Mean square	*F*	*p*	η_p^2
Number of attributes	1.018	2	0.509	153.127	0.000	0.012
Models	595.986	2	297.993	89683.310	0.000	0.881
Number of attributes* Models	52.177	4	13.044	3925.781	0.000	0.393
Error	80.712	24291	0.003			

Figure 1 shows that as the number of attributes increased, the ACRs of the DINO model and NPCD decreased, whereas the ACRs of the ANN increased. In addition, it was observed that the DINO model had the highest average ACR and that ANN had the lowest average ACR under all conditions of the number of attributes.

Figure 1

The ACRs of the DINO Model, ANN, and NPCD according to the number of attributes



The results of the PCRs of the DINO model, ANN, and NPCD, and the comparison of the number of attributes (3, 5, and 7) factor are presented in Table 2. Table 2 demonstrates that the interaction between the number of attributes of the PCRs and the models showed a statistically significant difference ($p<.01$) and that this interaction had a small effect (η_p^2 = .29).

Table 2

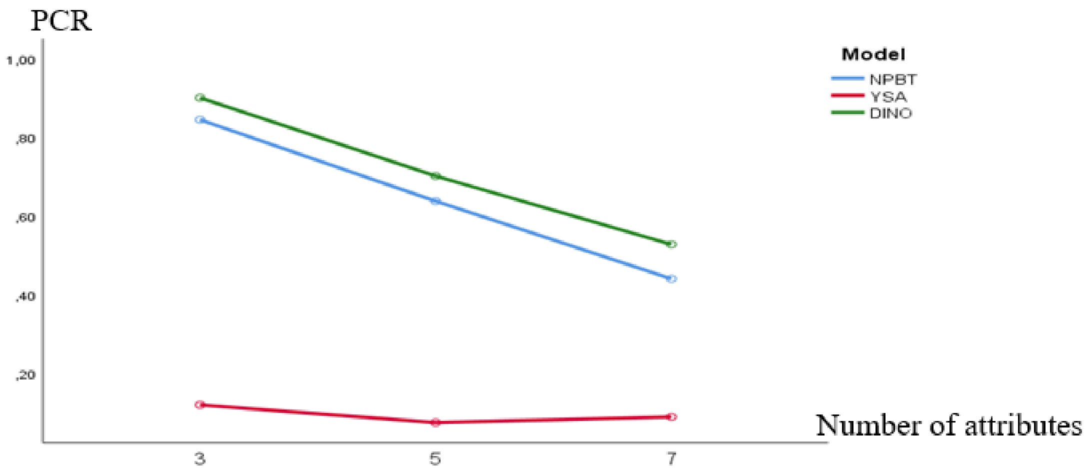
Factorial ANOVA Results of the DINO model, ANN, and NPCD's PCRs according to the Number of Attributes

Source of variation	Sum of squares	*df*	Mean square	*F*	*p*	η_p^2
Number of attributes	295.924	2	147.962	12421.698	0.000	0.506
Models	1847.983	2	923.991	77570.806	0.000	0.865
Number of attributes* Models	116.889	4	29.222	2453.265	0.000	0.288
Error	289.344	24291	0.012			

Figure 2 shows that the PCRs of the DINO model, ANN, and NPCD decreased as the number of attributes increased. It can also be seen that the DINO model had the highest average ACR and that ANN had the lowest average ACR under all conditions of the number of attributes.

Figure 2

The PCRs of the DINO Model, ANN, and NPCD according to the number of attributes



1.2. Comparison of the ACRs and PCRs of the DINO model, ANN, and NPCD according to the sample sizes (30, 100, and 500)

The results of the ACRs of the DINO model, ANN, and NPCD and the comparison of the sample sizes (30, 100, and 500) factor are presented in Table 3. Table 3 shows that the interaction between the sample sizes of the ACRs and the models was statistically significant (p<.01) and that this interaction had a small effect (η_p^2 = .002).

Table 3

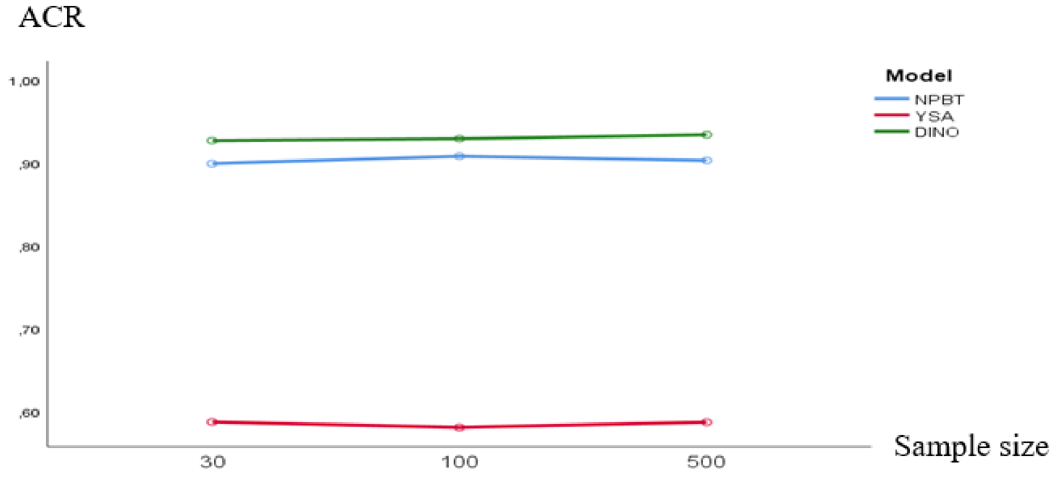
Factorial ANOVA Results of the DINO model, ANN, and NPCD's ACRs according to the Sample Sizes

Source of variation	Sum of squares	*df*	Mean square	*F*	*p*	η_p^2
Sample size	0.051	2	0.026	4.655	0.010	0.000
Models	595.986	2	297.993	54159.499	0.000	0.817
Sample size * Models	0.203	4	0.051	9.244	0.000	0.002
Error	133.652	24291	0.006			

Figure 3 shows that the ACRs of the NPCD, ANN, and DINO model did not change as the sample size increased. It can be seen that the DINO model had the highest average ACR and that ANN had the lowest average ACR under all conditions of the sample sizes.

Figure 3

ACRs of the DINO Model, ANN, and NPCD according to the sample sizes



The results of the PCRs of the DINO model, ANN, and NPCD and the comparison of the sample sizes (30, 100, and 500) factor are presented in Table 4. Table 4 shows that the interaction between the sample size of PCRs and the models showed a statistically significant difference (p<.01) and that this interaction had a small effect (η_p^2 = .003).

Table 4

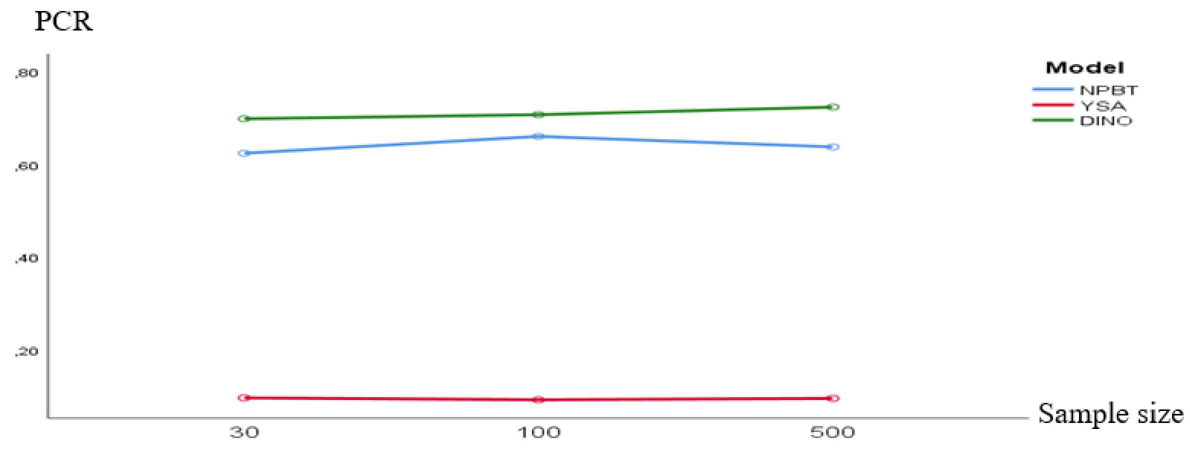
Factorial ANOVA Results of the DINO model, ANN, and NPCD's PCRs according to the Sample Sizes

Source of variation	Sum of squares	*df*	Mean square	*F*	*p*	η_p^2
Sample size	0.970	2	0.485	16.849	0.000	0.001
Models	1847.983	2	923.991	32094.844	0.000	0.725
Sample size* Models	1.864	4	0.466	16.188	0.000	0.003
Error	699.323	24291	0.029			

Figure 4 shows that when the sample sizes increased from 30 to 100, the PCRs of the DINO model and NPCD increased, whereas the PCRs of the ANN decreased. When the sample size was 500, it was observed that the PCRs of the NPCD decreased, the PCRs of the ANN showed no change, and the PCRs of the DINO model increased. Moreover, it was observed that the DINO model had the highest average PCR and that the ANN had the lowest average PCR under all conditions of the sample sizes.

Figure 4

The PCRs of the DINO Model, ANN, and NPCD according to sample sizes



1.3. Comparison of the ACRs and PCRs of the DINO model, ANN, and NPCD according to the number of items (15, 30, and 45)

The results of the ACRs of the DINO model, ANN, and NPCD and the comparison of the number of items (15, 30, and 45) factor are presented in Table 5. Table 5 shows that the interaction between the number of items of the ACRs and the models was statistically significant ($p<.01$) and that this interaction had a medium effect (η_p^2 = .38).

Table 5

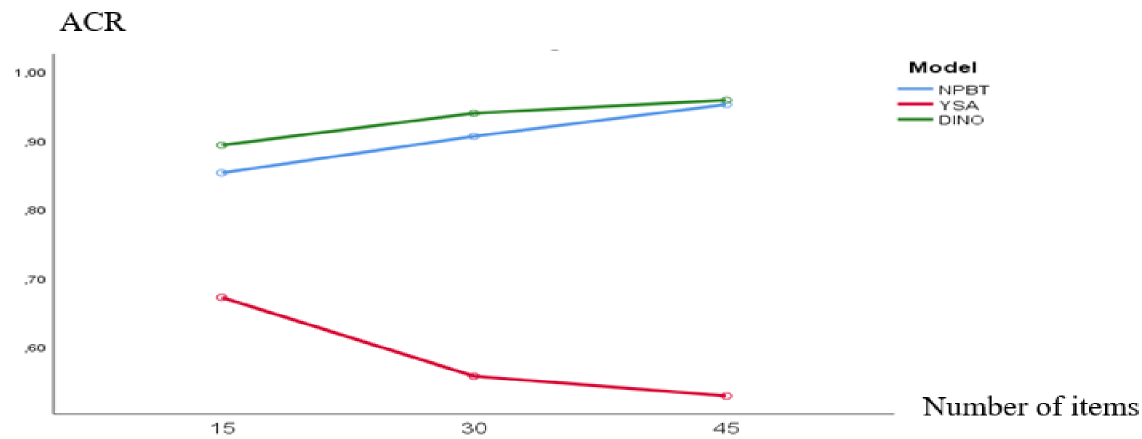
Factorial ANOVA Results of the DINO model, ANN, and NPCD's ACRs according to the Number of Items

Source of variation	Sum of squares	*df*	Mean square	*F*	*p*	η_p^2
Number of items	0.611	2	0.306	89.075	0.000	0.007
Models	595.986	2	297.993	86883.579	0.000	0.877
Number of items* Models	49.983	4	12.496	3643.283	0.000	0.375
Error	83.313	24291	0.003			

Figure 5 shows that as the number of items increased, the ACRs of the DINO model and NPCD increased, and the ACRs of the ANN decreased. When the number of items was 15 and 30, the DINO model had the highest average ACR, whereas the ANN had the lowest average ACR. Furthermore, when the number of items was 45, the DINO model and NPCD had the highest average ACR, whilst ANN had the lowest average ACR.

Figure 5

ACRs of the DINO Model, ANN, and NPCD according to the number of items



The results of the PCRs of the DINO model, ANN, and NPCD and the comparison of the number of items (15, 30, and 45) factor are presented in Table 6. Table 6 shows that the interaction between the number of items of the PCRs and the models was statistically significant (p<.01) and that this interaction had a small effect (η_p^2 = .19).

Table 6

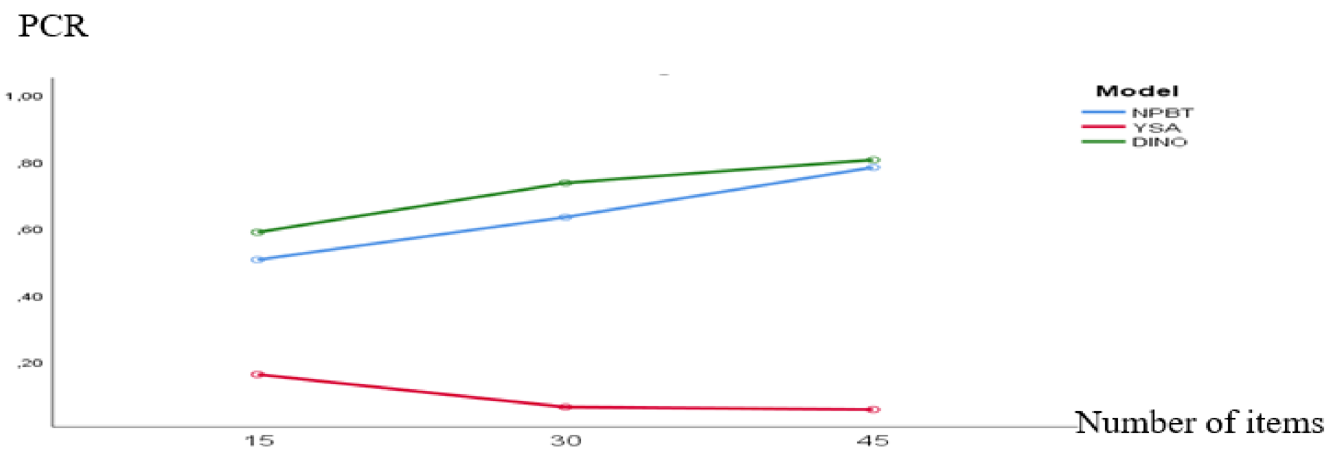
Factorial ANOVA Results of the DINO model, ANN, and NPCD's PCRs according to the Number of Items

Source of variation	Sum of squares	*df*	Mean square	*F*	*P*	η_p^2
Number of items	68.501	2	34.251	1621.845	0.000	0.118
Models	1847.983	2	923.991	43753.111	0.000	0.783
Number of items* Models	120.672	4	30.168	1428.522	0.000	0.190
Error	512.985	24291	0.021			

Figure 6 shows that when there was an increase in the number of items, the PCRs of the DINO model and NPCD increased, and the PCRs of the ANN decreased. It is also demonstrated that the DINO model had the highest average PCR and that ANN had the lowest average PCR under all conditions of the number of items.

Figure 6

PCRs of the DINO Model, ANN, and NPCD according to the number of items



1.4. Comparison of the ACRs and PCRs of the DINO Model, ANN, and NPCD according to the Missing Data Rates (0, .5, and .10)

The results of the ACRs of the DINO model, ANN, and NPCD and the comparison of the missing data rate (0, .05, and .10) factor are presented in Table 7. Table 7 shows that the interaction between the missing data rate and the models was statistically significant (p<.01) and that this interaction had a small effect (η_p^2 = .001).

Table 7

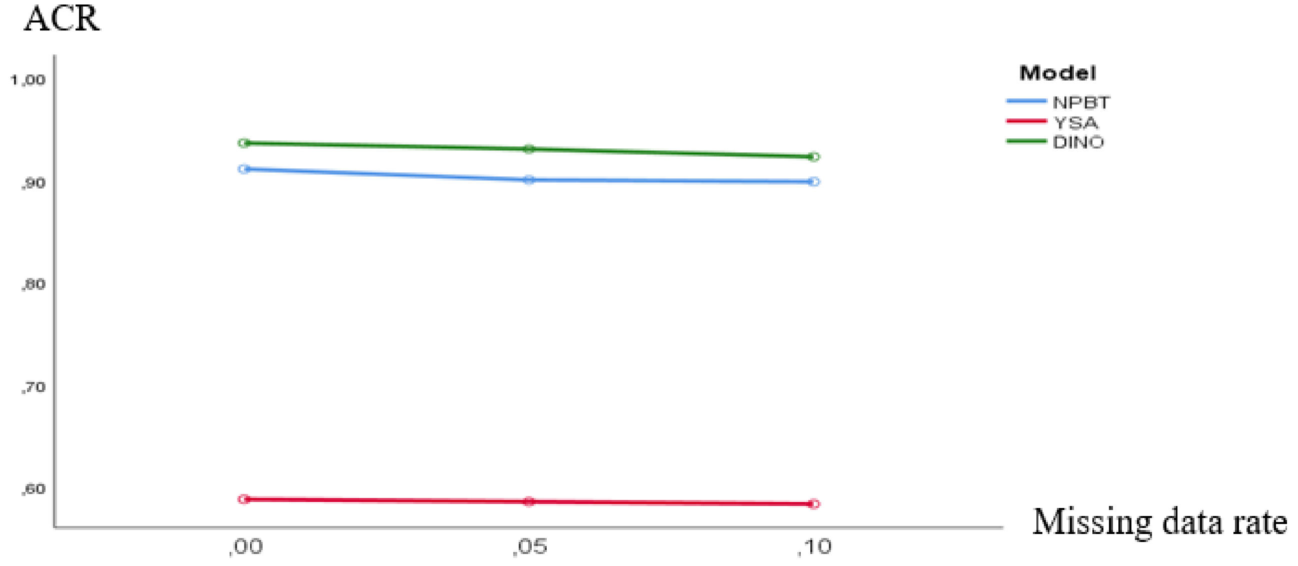
Factorial ANOVA Results of the DINO model, ANN, and NPCD's ACRs according to the Missing Data Rate

Source of variation	Sum of squares	*df*	Mean square	*F*	*p*	η_p^2
Missing data rate	0.424	2	0.212	38.614	0.000	0.003
Models	595.986	2	297.993	54266.062	0.000	0.817
Missing data rate * Models	0.093	4	0.023	4.235	0.002	0.001
Error	133.390	24291	0.005			

Figure 7 shows that the DINO model had the highest average ACRs compared with the ANN and NPCD, whereas the ANN had the lowest average ACRs under all conditions of the missing data rate.

Figure 7

The ACRs of the DINO Model, ANN, and NPCD according to the missing data rate



The results of the PCRs of the DINO model, ANN, and NPCD and the comparison of the missing data rate (0, .05, and .10) factor are presented in Table 8. Table 8 shows that the interaction between the missing data rate of the PCRs and the models was statistically significant ($p<.01$) and that this interaction had a small effect (η_p^2 = .001).

Table 8

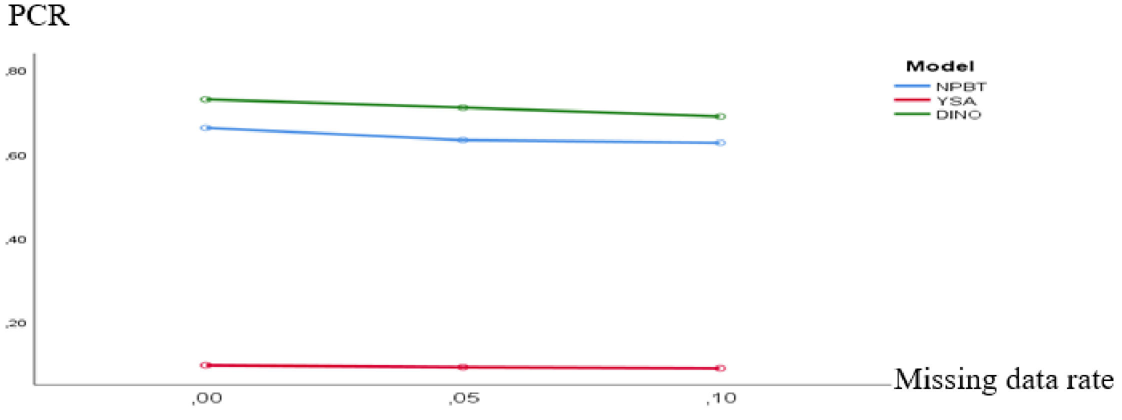
Factorial ANOVA Results of the DINO model, ANN, and NPCD's PCRs according to the Missing Data Rate

Source of variation	Sum of squares	*df*	Mean square	*F*	*p*	η_p^2
Missing data rate	3.230	2	1.615	56.219	0.000	0.005
Models	1847.983	2	923.991	32160.645	0.000	0.726
Missing data rate * Models	1.035	4	0.259	9.004	0.000	0.001
Error	697.893	24291	0.029			

Figure 8 shows that the DINO model had the highest PCRs compared with ANN and NPCD and that ANN had the lowest PCRs under all conditions of the missing data rate.

Figure 8

The PCRs of the DINO Model, ANN, and, NPCD according to the missing data rate



As a result of the factorial ANOVAs, the interaction effects for all conditions were found to be statistically significant. In all conditions, NPCD had slightly lower but comparable classification rates than the DINO model, while ANN always had lower rates than the NPCD and DINO model.

2. Similarities of the classification rates of the DINO model, ANN, and NPCD in the real dataset according to the number of attributes (3, 7, and 11) and sample sizes (30, 100, and 500)

2.1. Similarity of the classification rates of the DINO model, ANN, and NPCD according to the number of attributes (3, 7, and 11)

The SACRs and SPCRs of the DINA model,ANN, and NPCD based on different numbers of attributes (3,7, and 11) is presented in Table 9.

Table 9

Variation of SACR and SPCRs of the DINO Model, ANN, and NPCD according to the Number of Attributes in the Real Dataset

		Result of the Q matrix based on the technical reports						Result of the validated Q matrix					
		NPCD-ANN		NPCD-DINO		ANN-DINO		NPCD-ANN		NPCD-DINO		ANN-DINO	
N	*A*	SACR	SPCR	SACR	SPCR	SACR	SPCR	SACR	SPCR	SACR	SPCR	SACR	SPCR
30	3	.730	.471	.941	.875	.705	.472	.944	.867	.822	.533	.789	.467
	7	.714	.201	.808	.302	.721	.207	.786	.233	.771	.400	.805	.267
	11	.570	.002	.827	.070	.573	.003	.800	.033	.854	.267	.624	.00
100	3	.773	.532	.843	.684	.722	.481	.783	.470	.867	.710	.717	.380
	7	.691	.127	.776	.316	.658	.062	.741	.330	.771	.430	.819	.410
	11	.706	.038	.805	.087	.685	.014	.765	.090	.805	.080	.722	.080
500	3	.781	.515	.879	.664	.696	.426	.717	.436	.861	.668	.761	.496
	7	.682	.141	.777	.326	.679	.157	.790	.362	.860	.594	.824	.418
	11	.654	.026	.833	.123	.624	.018	.637	.030	.852	.334	.612	.018

Note: *A: Number of attributes, N: Sample size*

The SACRs and SPCRs obtained from the validated Q matrix were higher than the SACRs and SPCRs obtained from the Q matrix based on technical reports. In the findings obtained from Q matrices, the similarity between the SACRs and SPCRs of the DINO model and NPCD was high. It can be stated that as the number of attributes increased in the Q matrices, the SACRs between the ANN and NPCD and the SACRs between the NPCD and DINO model first decreased and then increased, whereas the SPCRs generally decreased.

2.2. Similarity of the classification rates of the DINO model, ANN, and NPCD according to the sample sizes (30, 100, and 500)

The SACRs and SPCRs of the DINO model, ANN, and NPCD according to different sample sizes (30, 100, and 500) are examined in Table 10.

Table 10

Variation of SACR and SPCRs of the DINO models, ANN and NPCD according to the Sample Sizes in the Real Dataset

		Result of the Q matrix based on the technical reports						Result of the validated Q matrix					
		NPCD-ANN		NPCD-DINO		ANN-DINO		NPCD-ANN		NPCD-DINO		ANN-DINO	
N	*A*	SACR	SPCR	SACR	SPCR	SACR	SPCR	SACR	SPCR	SACR	SPCR	SACR	SPCR
3	30	.730	.471	.941	.875	.705	.472	.944	.867	.822	.533	.789	.467
	100	.773	.532	.843	.684	.722	.481	.783	.470	.867	.710	.717	.380
	500	.781	.515	.879	.664	.696	.426	.717	.436	.861	.668	.761	.496
7	30	.714	.201	.808	.302	.721	.207	.786	.233	.771	.400	.805	.267
	100	.691	.127	.776	.316	.658	.062	.741	.330	.771	.430	.819	.410
	500	.682	.141	.777	.326	.679	.157	.790	.362	.860	.594	.824	.418
11	30	.570	.002	.827	.070	.573	.003	.800	.033	.854	.267	.624	.00
	100	.706	.038	.805	.087	.685	.014	.765	.090	.805	.080	.722	.080
	500	.654	.026	.833	.123	.624	.018	.637	.030	.852	.334	.612	.018

Note: A: Number of attributes, N: Sample size

This table shows that the SACRs and SPCRs obtained from the validated Q matrix were higher than the SACRs obtained from the Q matrix based on the technical reports. In the findings obtained from the Q matrices, no consistency was found regarding the increase or decrease of the SACR and SPCR values of NPCD-ANN, the NPCD-DINO model, and ANN-DINO model. In addition, it was observed that the SACRs and SPCRs between the DINO model and NPCD were more similar than those between ANN and NPCD and those between ANN and the DINO model in the findings obtained from the Q matrices.

Discussion

The aims of the study were to compare the attribute and pattern-level classification rates of the DINO model, ANN, and NPCD on the DINO-based simulation datasets based on various conditions and to examine the similarities between the classification rates of the DINO model, ANN, and NPCD on the PISA 2015 CPS dataset. In the current study, simulation datasets were generated similar to the structure of the real datasets in order to obtain comparable results from both datasets. With these aims in mind, the structure of the PISA 2015 CPS competency was examined, and it was found that there was no sequential or prerequisite relationship among the attributes, namely the problem-solving skills and collaboration skills, which constitute the PISA 2015 CPS competency. In other words, a student who has one or more problem-solving skills can solve a problem even if s/he does not have the other skills (Yavuz, 2014), which indicates that these skills are compensatory. Since all of the items are not shared in the technical reports on the PISA 2015 CPS competency, the extent to which each attribute contributes to the correct response cannot be determined for the items that require more than one attribute for the correct response. Therefore, the DINO model, which assumes that each attribute contributes equally to the correct response, was selected for the analysis of CPS competency and the generation of simulation datasets. In addition, it was found that one of the skills of the CPS competency, which consists of 12 skills, was not measured in the PISA 2015 CPS administration (see OECD, 2017). For this reason, the conditions of the attribute factor in the research were set as 3, 7, and 11. Moreover, since some conditions prevented the modeling, model-based data imputation methods could not be used to impute missing data. Instead, a two-way data imputation method was used in the current study. Therefore, the findings of this study were discussed within these limitations.

Discussion Of The Simulation Dataset

This study differs from other studies with ANN (see Cui, et al., 2016; Guo et al., 2017; McCoy & Willse, 2014; Paulsen, 2019; Paulsen & Valdivia, 2022; Shu et al., 2013) due to the DINO-based generation of simulation datasets. All the findings were discussed in light of the findings of previous studies conducted with the DINO and DINA models. As a result of the analyses, it was found that ANN always had lower rates than the DINO model and NPCD, whereas NPCD had slightly lower but comparable classification rates than the DINO model in all conditions. In parallel with this result, McCoy and Willse (2014) and Paulsen (2019), who studied the classification rates of the DINA model, ANN, and NPCD, found that the classification rates of NPCD and the DINA model were similar to each other whilst ANN consistently had lower classification rates in comparison with the DINA model and NPCD.

Shu et al. (2013) compared the classification rates of ANN, MCMC-ANN, JMLE-ANN, and the DINA model in simulation datasets with low item discrimination (s and g parameter values between .2 and .4) and a complex Q matrix structure. Shu et al. (2013) found that the DINA model cannot make estimations when the sample size is 50 or less in cases with four attributes and when the sample size is 150 or less in cases with six attributes. The studies in the related literature (Chiu et al., 2017; Paulsen, 2019; Paulsen & Valdivia, 2022) have also shown that the DINA model can also make estimations in small samples of 25 and 30 participants thanks to the increase in computational capacity and the use of the EM algorithm in parameter estimation of items. Likewise, it was also found in the current study that the DINO model could make estimations in the conditions of 30 participants, the smallest sample size.

In addition to the increase in the computational capacity and the use of the EM algorithm, the DINO model is thought to have higher classification rates than ANN and NPCD in small samples due to the fact that the simulation datasets were generated based on the DINO model and the balanced distribution of attributes and items constituting the Q matrix structures was taken into consideration in the data generation (see de la Torre, 2008; de la Torre et al., 2010; Rupp & Templin, 2008a). In the literature, there are studies with various results. In parallel with the results of the current study, Chiu and Douglas (2013) stated that the maximum likelihood estimation estimates better with the most suitable parametric model for the data structure; therefore, the DINO model can make better estimations than NPCD in some small DINO-based simulation samples. Similar to the results of the DINO model and NPCD, Cui et al. (2016) and Shu et al. (2013) also found that the DINA model had higher classification rates than ANN in DINA-based simulation datasets.

In the literature, there are studies showing that ANN and NPCD have higher and lower classification rates than the DINO or DINA models, depending on the conditions. For instance, McCoy and Willse (2014) studied DINA-based datasets and found that the classification rates of the DINA model were slightly higher than NPCD and considerably higher than ANN in some conditions. Chiu et al. (2017) stated that NPCD had higher classification rates than the DINA model when the sample size was 100 or less, whereas the DINA model had higher classification rates than NPCD when the sample size was 500. Furthermore, Paulsen (2019) found that NPCD had better classification rates than the DINA model when the sample size was 25, the number of items was small, and the item discrimination was low. Finally, Ma et al. (2020) found that NPCD had higher classification rates than the DINA model in samples of 30 and 50 and that the DINA model had higher classification rates than NPCD in samples of 200 and 500.

There are also studies in the literature showing that ANN and NPCD have higher classification rates than the DINA and DINO models in all conditions. For instance, Akbay (2016) examined the classification rates of the NPCD, DINO, and DINA models in various conditions by generating simulation datasets of 250, 500, and 1000 participants based on DINA and DINO. He found in his study that NPCD had higher classification rates than the DINO model in a DINO-based dataset, and NPCD had higher classification rates than the DINA model for a DINA-based dataset. In their DINA-based dataset with a sample of 5000 participants, Guo et al. (2017) found that ANN had a higher classification rate than the DINA model.

Discussion of the real dataset (PISA 2015 CPS)

In the current study, no consistency was found regarding the increase, decrease, or stability of the SACRs and SPCRs of the DINO model, ANN, and NPCD when the sample size increased (30, 100, and 500). This is an expected finding, considering that, in theory, the ACRs and PCRs of ANN and NPCD are not affected by sample size. No study was found in the literature that had investigated either the similarity of classification rates of the DINO model, ANN, and NPCD or the similarity of classification rates of the DINA model, ANN, and NPCD in the real datasets. Nevertheless, only three studies were found in the literature which estimated the similarity of classification rates of different models. Chiu and Douglas (2013) calculated the similarity of classification rates of NPCD and HODINA, while Chiu et al. (2017) calculated the similarity of classification rates of GNPCD and G-DINA. Lim and Drasgow (2017) calculated the similarity of classification rates of NPCD based on expert opinion. However, these studies were carried out with a single sample.

In conclusion, the similarity between the classification rates of the DINO model and NPCD was observed across both simulation and real datasets. In addition, no consistency was found regarding the increase or decrease in the classification rates of ANN and NPCD in simulation datasets and the similarities of these rates in real datasets when the sample size increased in both datasets. It was also observed that the variations in the classification rates and the similarity of these rates differed as the number of attributes in the simulation and the real datasets differed.

There are some limitations in this study. First, since not all of the real data sets were shared, it could not be determined which attribute contributed more to the correct answer of the items. Therefore, the DINO model, which is assumed to contribute equally to each attribute in answering the item correctly, was chosen as a parametric analysis. Second, the first Q matrices were created according to the technical reports (OECD, 2017a; 2017b) based on this reason. Third, model-based data imputation methods could not be used in the missing data imputation since some conditions prevented the establishment of the model; instead, the two-way data assignment method was used. There are some suggestions considering these limitations.

The findings of the current study have shown that the classification rates of the DINO model and NPCD were similar. It is thought that evaluating the results of the DINO model and NPCD together will increase the classification reliability and hence can contribute to the reliable assessments of students as well as their placement into correct attribute classes. For this reason, it is suggested that implementers use the DINO model and NPCD together if they are going to perform cognitive diagnoses in small samples.

The classification rates of the DINO model, ANN, and NPCD can be further investigated by changing the attributes, research factors, software, and packages of the simulation datasets used in this study. Based on the analysis of the classification rates of the DINO model, ANN, and NPCD, examining the PISA 2015 CPS construct with a three-attribute Q matrix was found to be more reliable. In OECD (2017a; 2017b) technical reports, these three attributes are described as competency areas of the "establishment and maintenance of a common understanding", the "identification of proper actions to solve the problem", and the "establishment and maintenance of team organization". Researchers who intend to investigate the PISA 2015 CPS construct by using different methods are recommended to conduct their investigation based on these three sub-competency areas.

Acknowledgments

This study was supported by the Scientific and Technological Research Council of Turkey under Grant 2228-B.

Declarations

Conflict of Interest: The authors of the article declare that they have no conflict of interest with any person or organization that may be a party to this study.

Ethical Approval: Simulated and open-access data were used in this study. Therefore, ethical approval is not required.

References

Akbay, L. (2016). Relative efficiency of the nonparametric approach on attribute classification for small sample cases. *Journal of European Education 6*(1), 43-59.

Chiu, C.-Y., & Douglas, J. (2013). A nonparametric approach to cognitive diagnosis by proximity to ideal response patterns. *Journal of Classification*, 30, 225-250. doi:10.1007/s00357-013-9132-9

Chiu, C.-Y., Douglas, J., & Li, X. (2009). Cluster analysis for cognitive diagnosis: Theory and applications. *Psychometrika, 74*(4), 633-665.

Chiu, C.-Y., & Köhn, H.-F. (2019). Consistency theory for the general nonparametric classification method. *Psychometrika, 84*(3), 830-845.

Chiu, C.-Y., Sun, Y., & Bian, Y. (2017). Cognitive diagnosis for small educational programs: The general nonparametric classification method. *Psychometrika*, *83*(2), 355-375.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Erlbaum.

Cui, Y., Gierl, M., & Guo, Q. (2016). Statistical classification for cognitive diagnostic assessment: An artificial neural network approach. *Educational Psychology: An International Journal of Experimental Educational Psychology, 36*(6), 1065-1082.

Cui, Y., Gierl, M., & Guo, Q. (2017). The rule space and attribute hierarchy methods. In A. A. Rupp & J. P. Leighton (Eds.), *The handbook of cognition and assessment frameworks, methodologies, and applications* (pp. 354-378). John Wiley & Sons.

Dai, S. (2017). *Investigation of missing responses in implementation of cognitive diagnostic models.* (Doctoral dissertation). Indiana University, Boston.

Dai, S., Wang, X., & Svetina, D. (2019, March). Test Data Imputation [software package in R]. https://cran.r-project.org/web/packages/TestDataImputation/index.html

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., & Chiu, C.-Y. (2015). A general method of empirical Q-matrix validation. *Psychometrika*, *81*(2), 253-273.

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.

Fritsch, S., Guenther, F., Wright, M. N., Suling, M., & Mueller, S. M. (2019, February). Neuralnet [software package in R]. https://cran.r-project.org/web/packages/neuralnet/neuralnet.pdf

Garson, G. D. (1998). *Neural networks: An introductory guide for social scientists*. Sage.

Gierl, M. J., Cui, Y., & Hunka, S. (2007, April). *Using connectionist models to evaluate examinees' response patterns on tests.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, Chicago, IL, USA.

Gierl, M. J., Cui, Y., & Hunka, S. (2008). Using connectionist models to evaluate examinees' response patterns on tests. *Journal of Modern Applied Statistical Methods*, 7, 234-245. doi:10.22237/jmasm/1209615480

Guo, Q., Cutumisu, M., Cui, Y. (2017). *A neural network approach to estimate student skill mastery in cognitive diagnostic assessments*. Poster presented at the 10th International Conference on Educational Data Mining, Wuhan, Hubei Province, in China.

He, Q., von Davier, M., Greiff, S., Steinhauer, E. W., & Borysewicz, P. B. (2017). Collaborative problem-solving measures in the Programme for International Student Assessment (PISA). In A. A. von Davier, M. Zhu & P. C. Kyllonen (Eds.), *Methodology of educational measurement and assessment: Innovative assessment of collaboration* (pp. 95-112). Springer.

IBM Corp. (2019). IBM SPSS Statistics for Windows (Version 22.0) [Computer software]. https://www.ibm.com/tr-tr/analytics/spss-statistics-software

Lim, Y. S., & Drasgow, F. (2017): Nonparametric calibration of ıtem-by-attribute matrix in cognitive diagnosis. *Multivariate Behavioral Research, 52*(5), 562-575.

Ma, C., de la Torre, J., & Xu, G. (2020). *Bridging Parametric and Nonparametric Methods in Cognitive Diagnosis.* Retrieved from arXiv:2006.15409

Ma, W., de la Torre, J., Sorrel, M. & Jiang, Z. (2020, May). GDINA [software package in R]. https://cran.r-project.org/web/packages/GDINA/index.html

McCoy, T., & Willse, J. (2014, April). *Accuracy of neural network versus nonparametric approaches in diagnostic classification.* Paper presented at the National Council on Measurement in Education, Washington, DC, USA.

No Child Left Behind Act of 2001, Pub. L. No. 107-110. Retrieved from http://thomas.loc.gov/

OECD (2017a). *PISA 2015 collaborative problem-solving framework.* France. https://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf

OECD (2017b). *PISA 2015 results (Volume V): Collaborative problem solving.* France. https://www.oecd.org/publications/pisa-2015-results-volume-v-9789264285521-en.htm#:~:text=PISA%202015%20Results%20(Volume%20V)%3A%20Collaborative%20Problem%20Solving%2C%20is,try%20to%20solve%20a%20problem

Paulsen, J. (2019). *Examining cognitive diagnostic modeling in small sample contexts*. (Doctoral dissertation). Indiana University, Boston.

Paulsen, J. & Valdivia, D. S. (2022). Examining cognitive diagnostic modeling in classroom assessment conditions. *Journal of Experimental Education, 90*(4), 916-933. doi:10.1080/00220973.2021.1891008

Robitzsch, A., Kiefer, T., George, A. C., & Uenlue, A. (2019, September). CDM [software package in R]. https://cran.r-project.org/web/packages/CDM/index.html

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

Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement*. Guilford.

Shu, Z., Henson, R., & Willse, J. (2013). *Using neural network analysis to define methods of DINA model estimation for small sample sizes*. *Journal of Classification*, 30, 173-194. doi:10.1007/s00357-013-9134-7

Shuying, S. (2016). *Nonparametric diagnostic classification analysis for testlet based tests.* (Doctoral dissertation). The University of North Carolina at Greensboro, USA.

Stekhoven, D. J. (2016, August). missForest [software package in R]. https://cran.r-project.org/web/packages/missForest/missForest.pdf

Sünbül, S. (2018). The impact of different missing data handling methods on DINA model. *International Journal of Evaluation and Research in Education, 7*(1), 77-86.

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods, 11*(3), 287-305.

Wang, M.-J. (2015). *Computer-based assessment of collaborative problem-solving for intermediate elementary students with one on one, human-to-agent approach.* (Master's dissertation). National Taichung University, Taiwan.

Wang, S., & Douglas, J. (2015). Consistency of nonparametric classification in cognitive diagnosis. *Psychometrika, 80*(1), 85-100.

Yavuz, E. (2014). *Determining the problem solving process skills of the primary education pre-service mathematics teachers as defined in PISA*. Gazi University, Ankara.

Zheng, Y., Chiu, C.-Y., & Douglas, J. A. (2019, November). NPCD [software package in R]. https://cran.r-project.org/web/packages/NPCD/NPCD.pd

ISSN: 1309 – 6575
Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi
Journal of Measurement and Evaluation in Education and Psychology
Research Article; 2023; 14(4); 413-439

Appendices

Table 11

Q Matrices Used in Generating Simulation Datasets

Item no	The Q matrix with three attributes			The Q matrix with five attributes					The Q matrix with seven attributes						
1	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0
2	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0
3	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0
4	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0
5	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0
6	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0
7	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1
8	1	1	0	1	1	0	0	0	1	1	0	0	0	0	0
9	0	1	1	0	0	1	1	0	0	1	1	0	0	0	0
10	1	0	1	1	1	0	0	0	0	0	1	1	0	0	0
11	0	1	1	0	1	1	0	0	0	0	0	1	1	0	0

*Asst. Prof. Dr.,Erciyes University, Faculty of Education, Kayseri-Türkiye, emineyavuz@erciyes.edu.tr, ORCID: 0000-0002-1991-1416
** Prof. Dr., Gazi University, Faculty of Education, Ankara-Türkiye, hakanyavuzatar@gmail.com, ORCID ID: 0000-0001-5372-1926

To cite this article:
Yavuz, E. & Atar, H.,Y. (2023). A Comparison of the classification performances of the DINO model, artificial neural networks and non-parametric cognitive diagnosis. *Journal of Measurement and Evaluation in Education and Psychology, 14*(4), 413-439. https://doi.org/10.21031/epod.1246752

Received: 2.02.2023
Accepted: 20.10.2023

12	1	0	1	0	0	1	1	0	0	0	0	0	1	1	0
13	1	1	0	0	0	0	1	1	0	0	0	0	0	1	1
14	1	1	1	1	0	0	1	1	1	0	0	1	0	0	1
15	1	1	1	1	0	1	0	1	1	1	1	0	0	0	0

Table 12

Q Matrices for Top 17 Items for Real Data

Item no	The Q matrix with three attributes			The Q matrix with seven attributes							The Q matrix with eleven attributes										
	1	2	3	A	B	C	D	1	2	3	A1	A2	B1	B2	B3	C1	C2	C3	D1	D2	D3
1	0	1	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
2	1	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
3	0	1	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
4	1	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
5	1	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
6	1	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
7	0	0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0

8	1	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
9	1	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
10	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1
11	0	1	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
12	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
13	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0
14	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
15	1	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
16	1	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
17	1	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0

Table 13

The ACRs and PCRs of NPCD, ANN, and the DINO Model under Study Conditions

			Classification Rates of NPCD						Classification Rates of ANN						Classification Rates of DINO					
			ACRs			PCRs			ACRs			PCRs			ACRs			PCRs		
			Missing Date Rate			Missing Date Rate			Missing Date Rate			Missing Date Rate			Missing Date Rate			Missing Date Rate		
A	*N*	*I*	0%	5%	10%	0%	5%	10%	0%	5%	10%	0%	5%	10%	0%	5%	10%	0%	5%	10%
3	30	15	.930	.903	.883	.810	.751	.707	.622	.580	.577	.214	.178	.167	.945	.933	.928	.850	.819	.806
		30	.956	.949	.941	.879	.862	.84	.429	.513	.496	.074	.093	.111	.978	.979	.967	.934	.938	.900
		45	.978	.979	.979	.938	.939	.943	.480	.476	.467	.107	.111	.092	.989	.984	.969	.966	.951	.911
	100	15	.877	.875	.902	.708	.699	.752	.508	.579	.576	.115	.168	.160	.941	.933	.926	.84	.818	.799
		30	.957	.949	.954	.883	.861	.873	.56	.447	.429	.161	.065	.073	.981	.983	.972	.948	.953	.920
		45	.971	.983	.968	.919	.950	.910	.484	.478	.475	.108	.104	.099	.989	.987	.978	.966	.962	.937
	500	15	.907	.869	.924	.761	.692	.801	.575	.512	.574	.163	.114	.162	.947	.940	.929	.852	.833	.805
		30	.945	.954	.944	.854	.875	.849	.430	.530	.432	.076	.115	.073	.974	.974	.968	.925	.927	.910
		45	.992	.982	.97	.977	.947	.915	.472	.482	.479	.100	.099	.101	.988	.985	.982	.963	.956	.948
5	30	15	.842	.796	.847	.471	.380	.477	.654	.684	.681	.127	.165	.148	.895	.885	.880	.588	.568	.539

		30	.883	.885	.899	.573	.572	.608	.536	.549	.591	.040	.044	.072	.936	.928	.926	.728	.704	.703
		45	.946	.948	.944	.775	.783	.765	.523	.484	.482	.041	.024	.022	.957	.943	.961	.803	.746	.829
	100	15	.881	.862	.853	.554	.526	.510	.668	.644	.654	.136	.119	.127	.886	.877	.869	.562	.527	.521
		30	.922	.91	.901	.686	.665	.619	.540	.540	.524	.044	.041	.033	.943	.939	.928	.764	.745	.728
		45	.977	.952	.943	.897	.794	.767	.488	.487	.485	.025	.025	.028	.967	.954	.947	.848	.801	.777
	500	15	.878	.849	.876	.560	.478	.552	.680	.699	.688	.152	.169	.156	.901	.891	.882	.607	.586	.559
		30	.909	.917	.898	.643	.684	.621	.535	.524	.595	.041	.036	.069	.944	.942	.931	.765	.756	.718
		45	.961	.940	.936	.828	.757	.742	.522	.514	.511	.041	.040	.037	.965	.961	.955	.847	.828	.809
7	30	15	.816	.818	.808	.285	.286	.288	.788	.761	.757	.193	.160	.152	.854	.854	.843	.375	.367	.347
		30	.844	.878	.87	.357	.446	.413	.632	.654	.629	.046	.048	.039	.909	.904	.892	.557	.525	.484
		45	.923	.925	.92	.606	.610	.581	.617	.565	.644	.036	.012	.052	.938	.934	.929	.665	.648	.620
	100	15	.823	.838	.793	.306	.358	.260	.795	.782	.757	.214	.188	.156	.868	.863	.979	.393	.388	.356
		30	.907	.874	.835	.538	.415	.355	.662	.668	.661	.058	.060	.057	.917	.909	.900	.575	.547	.517
		45	.957	.936	.929	.758	.662	.631	.570	.679	.562	.020	.077	.015	.940	.933	.922	.674	.647	.602
	500	15	.805	.777	.797	.271	.234	.257	.793	.786	.760	.203	.195	.158	.874	.865	.857	.415	.395	.38
		30	.894	.858	.832	.495	.384	.336	.670	.628	.636	.061	.038	.041	.923	.916	.907	.603	.579	.541
		45	.930	.916	.930	.630	.565	.630	.643	.566	.633	.051	.017	.044	.949	.944	.937	.712	.688	.658

Note: A: Number of attributes. N: Sample size. I: Number of items