Center for Open Access in Science ▪ https://www.centerprode.com/ojer.html
Open Journal for Educational Research, 2023, *7*(1), 53-62.
ISSN (Online) 2560-5313 ▪ https://doi.org/10.32591/coas.ojer.0701.05053k

Standard Setting with Artificial Neural Networks: TIMSS 2015 Mathematics Case

Mahmut Sami Koyuncu

Afyon Kocatepe University, Faculty of Education, Afyonkarahisar, TURKEY

Received: 3 April 2023 ▪ Revised: 20 July 2023 ▪ Accepted: 30 July 2023

Abstract

This study aims to demonstrate the optimal way to determine the cut-off score to be used to interpret the total scores obtained from an achievement test or scale using the Artificial Neural Networks method. To this end, the multiple-choice item responses in the Booklet-11 Mathematics subtest at the 8th grade level in the TIMSS 2015 Turkey sample dataset were used to determine the cut-off score for the achievement test. The item responses in the "Students Like Learning Mathematics Scale" in the TIMSS 2015 8th grade Mathematics Student Questionnaire were used to determine the cut-off score for the scale. The data were accessed from the TIMSS international database and the data were analyzed in MATLAB R2017b software. As a result of the study, the most appropriate cut-off score to be used for the evaluation of the total scores obtained from the TIMSS 2015 8th grade level Booklet-11 Mathematics subtest was determined as 45.5 out of 0-100 points with the Artificial Neural Network analysis method. The overall level of agreement between the cut-off score and the pass/fail classification based on 400 points, which is the lowest level of the TIMSS International Benchmark, was determined as 81%. The most appropriate cut-off score to be used for the evaluation of the scores obtained from the Students Like Learning Mathematics Scale (SLLSS) in the TIMSS 2015 8th grade student survey was determined as 19.6 out of 9-36 points. The overall level of agreement between the cut-off score and the classification of students who like/don't like learning mathematics using the criterion based on the expression given in the original scale description was found to be 83%. The results concluded that the validity of the standard-setting studies conducted with the artificial neural network method was high. As a result, researchers are recommended to use the Artificial Neural Networks method to determine the cut-off score to be used in the interpretation of the total scores obtained from the achievement test or the total scale scores obtained from the scales.

Keywords: artificial neural networks, standard setting, cut-off score, TIMSS 2015.

1. Introduction

It is essential to measure and evaluate achievement in education, recruitment, scientific research, and many other contexts. These measurements are a fundamental tool for assessing the performance of individuals or processes, making decisions, and monitoring progress. In measurement practices, the cut-off scores used to interpret the results of achievement tests or scales play a critical role. Cut-off score determination is also a standard-setting process. The literature hosts numerous standard setting methods. The current study addresses the Artificial Neural Networks (ANN) for standard setting.

Correspondence: Mahmut Sami Koyuncu, Afyon Kocatepe University, Faculty of Education, Afyonkarahisar, TURKEY.

ANNs are among data mining models as statistical classification techniques based on a predictive approach. ANN is an artificial network system inspired by the neural network structure of the brain. It is a mathematical model of brain activities (Shah & Murtaza, 2000). ANN is similar to the brain in that information is acquired by passing through a learning process and using the link power among neurons to store this information (Haykin, 1999). In this respect, the work of ANN is to gain learning, generalization, and recollection characteristics for the systems (Saraç, 2005). ANN can be used for nonlinear, multidimensional, complex, uncertain, missing, and error-prone data, especially when no mathematical model or algorithm exists for solving problems. ANN performs functions of prediction, classification, data association, data filtering, recognition and matching, diagnosis, and interpretation (Öztemel, 2003).

Artificial neural networks do not require assumptions regarding the distribution of data. In clustering studies, artificial neural networks can be employed instead of classical statistical methods. The most commonly used artificial neural networks in clustering studies are Self-Organizing Maps (SOM) neural networks. SOM networks are single-layer networks. SOM algorithm is indeed an unsupervised learning algorithm. The data to be used in the training of this network does not contain dependent variables. Often, these variables are referred to as features (Kohonen, 2001).

SOM networks are preferred for both clustering and visualization of data. These networks reduce multidimensional data into a two-dimensional map. SOM networks can fulfill the functions of both K-means and multidimensional scaling methods in classical statistics. That is, it both clusters and maps the elements in the data set. Therefore, these networks have become very popular in recent years (Bircan et al., 2010).

Despite the limited use of ANN in education, it appears to be widely used in transportation, medicine, biomedical industry, finance, stock exchange, and computer technology. However, research has revealed that ANN produces more accurate estimates and classification percentages than other regression and classification methods (Gorr et al., 1994; Ibrahim & Rusli, 2007; Subbanarasimha et al., 2000; Wilson et al., 1994;). Accordingly, the ANN analysis can be used as an alternative method in educational studies.

Scales and tests developed today are used in recruitment, education, choice of profession, decision-making about individuals, and clinical areas. Many researchers have problems with how and according to what to interpret the scores obtained from the scale/test they have developed. Within the scope of the current study, demonstrating how to determine the cut-off score based on the ANN method will help researchers in this regard. In addition, providing evidence on the validity of the cut-off score, which is neglected in many studies, increases the importance of the study.

2.1 *Purpose of the research*

This study aims to demonstrate how to determine the cut-off score to be used to interpret the total scores obtained from an achievement test or scale using the Artificial Neural Networks (ANN) method. The study also aims to examine the validity of the cut-off scores. To these ends, answers to the following questions were sought:

- What is the most appropriate cut-off score to be used to evaluate the total scores obtained from the TIMSS 2015 8th Grade Booklet-11 Mathematics subtest with the ANN analysis method?
- What is the distribution of students' achievement status according to the cut-off score determined by ANN for TIMSS international proficiency levels and Mathematics subtest?

- What is the most appropriate cut-off score to be used for the evaluation of the scores obtained from the "Students Like Learning Mathematics" scale in the TIMSS 2015 8th grade student questionnaire with the ANN analysis method?
- What is the accuracy between the cut-off scores of the Students Liking for Learning Mathematics Scale determined by TIMSS guidelines and the Students Like Learning Mathematics Scale determined by ANN?

2.2 *Method*

2.2.1 *Research model*

This is a descriptive study because it aims to show how to determine the cut-off score to be used to interpret the total scores obtained from an achievement test or scale with the Artificial Neural Network method and to examine the validity of the determined cut-off scores.

2.2.2 *Study group*

The research was performed with two different study groups. For the purpose of determining the cut-off score for the achievement test in the study, the data of 441 students in the TIMSS 2015 Turkey sample who took the Booklet-11 subtest of the 8th Grade Mathematics Test were used. Of these students, 50.6% (N=223) were female and 49.4% (N=218) were male. In the study, the data of 5,741 students in the TIMSS 2015 Turkey sample who answered the "Students Like Learning Mathematics Scale" at the 8th grade level were used to determine the cut-off score for the scale. Of these students, 49% (N=2,812) were female and 51% (2929) were male.

2.2.3 *Data description*

TIMSS, conducted every four years by the International Association for the Evaluation of Educational Achievement (IEA), also creates an international database that determines the trends in students' achievement in mathematics and science. The study data were obtained from the TIMSS 2015 international database (https://timssandpirls.bc.edu/timss2015/international-database/).

The 8th grade Booklet-11 mathematics subtest used in the study included 16 multiple-choice items with four options. In the Students Like Learning Mathematics Scale included in the TIMSS 2015 8th grade Mathematics Student Questionnaire, student scores are rated between 1-4 as 1=Disagree a lot, 2=Disagree a little, 3=Agree a little, and 4=Agree a lot and consists of a total of 9 items (Mullis et al., 2020).

2.2.4 *Data analysis*

The study basically serves two purposes. The first is to find the most appropriate cut-off scores for the mathematics achievement test and the Students Like Learning Mathematics Scale with ANN, and the second is to examine the validity of the cut-off scores.

After the data sets were obtained from the international database, the students' raw scores were obtained by giving 1 point for a correct answer and 0 points for an incorrect answer from the student responses in Booklet 11. The raw scores were then converted into a 100-point system, i.e., the maximum score was 100. After this process, the cut-off score was determined by using the SOM learning algorithm with the ANN analysis method. Similarly, for the Students Like

Learning Mathematics Scale, the students' responses were scored between 1-4, and the cut-off score was determined by the ANN analysis.

In the second stage, in order to provide evidence for the validity of the cut-off scores, the consistency between the classifications of students according to TIMSS 2015 international proficiency levels and the classifications made according to the cut-off score determined for the mathematics subtest was examined. TIMSS 2015 8th grade mathematics international proficiency levels are presented in Table 1 (Mullis et al., 2016).

Table 1. TIMSS 2015 international benchmarks of mathematics achievement

Score	**Benchmarks**
625	Advanced
550	High
475	Intermediate
400	Low
Below 400	Below Low

**: The level of students who do not even reach the lowest level in TIMSS*

In addition, the agreement between the pass/fail classification based on the lowest TIMSS International Benchmarks of 400 points and the pass/fail classification based on the cut-off score determined for the Booklet-11 mathematics subtest was examined.

The scores used in the evaluation of the scores from the Students Like Learning Mathematics Scale (SLLMS) are defined as follows (Mullis et al., 2020): "*Students Who Do Not Like Learning Mathematics had a score at or below the cut score corresponding to "disagreeing a little" with five of the nine statements and "agreeing a little" with the other four, on average. All other students Somewhat Like Learning Mathematics.*" Based on this definition, the criterion set for the original scale was based on 22 points (five items disagreeing a little, 5x2=10; four items agreeing a little, 4x3=12 Total=22). Students scoring 22 points and below were classified as not like learning mathematics, while those scoring above 22 points were classified as like learning mathematics. Then, the agreement between this classification and the classification based on the score determined by the ANN method was examined.

Sensitivity (true positive rate), Specificity (true positive rate), and Accuracy values were presented as agreement values. Data analysis was performed using the SPSS package program and MATLAB R2017b software.

2.3 *Results*

Within the scope of the study, firstly, the question "What is the most appropriate cut-off score to be used for the evaluation of the total scores obtained from the TIMSS 2015 8th Grade Booklet-11 Mathematics subtest by ANN analysis method?" was sought to be answered. Figure 1 shows the cut-off score determined by the ANN method for the pass/fail decision to be made according to the total score of the students in 16 multiple-choice mathematics subtests.

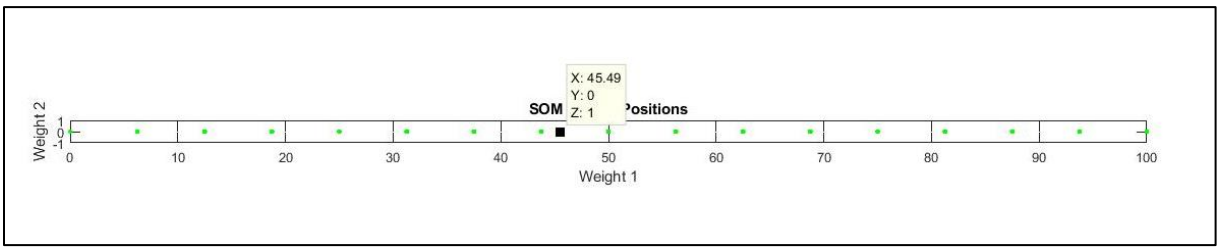

Figure 1. The cut-off score determined for the total score of the mathematics subtest (Booklet-11)

Figure 1 demonstrates that the most appropriate cut-off score to be used to evaluate the total scores obtained from the TIMSS 2015 8th grade level Booklet-11 Mathematics subtest by ANN analysis method was 45.49 out of 0-100 points. According to the cut-off score determined, the achievement status of 411 students in the study group in the mathematics test is shown in Figure 2.

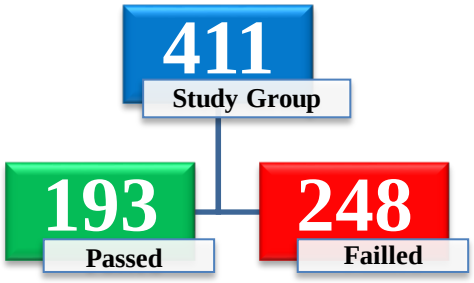

Figure 2. Mathematics subtest achievement status of the students in the study group

As given in Figure 2, according to the cut-off score determined, 193 students (47%) out of 411 students in the study were successful in the mathematics test, while 248 students (53%) were unsuccessful.

Within the scope of the study's second aim, the question "How is the distribution of the achievement status of the students according to the cut-off score determined by ANN for TIMSS international proficiency levels and Mathematics subtest?" was sought to be answered. Figure 3 shows the distribution of successful students according to TIMSS international proficiency levels and Figure 4 shows the distribution of unsuccessful students according to TIMSS international proficiency levels.

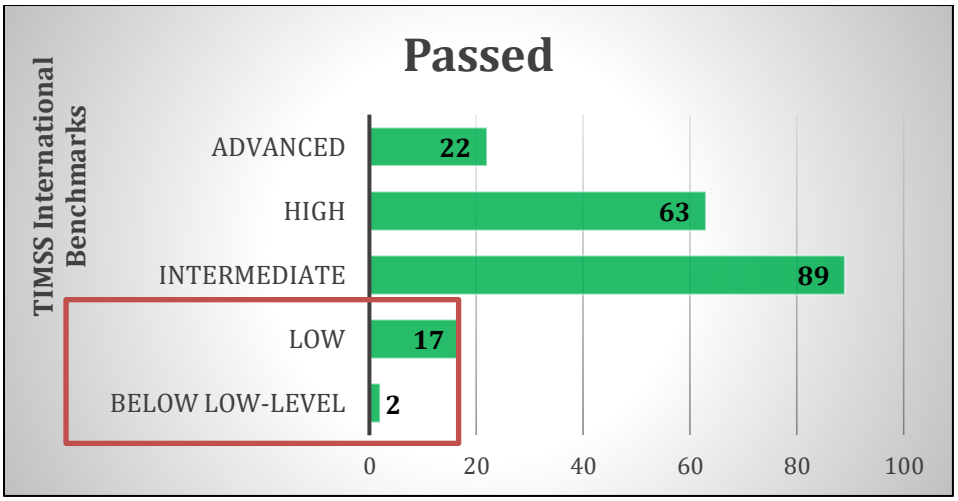

Figure 3. Distribution of the successful group according to TIMSS international benchmarks

When Figure 3 is analyzed, it is seen that 1.0% (N=2) of the 193 students in the successful group were below the Low Level, 8.8% (N=17) were at the Low Level, 46.1% (N=89) were at the Intermediate Level, 32.6% (N=63) were at the Upper Level and 11.4% (N=22) were at the Advanced TIMSS international proficiency level.

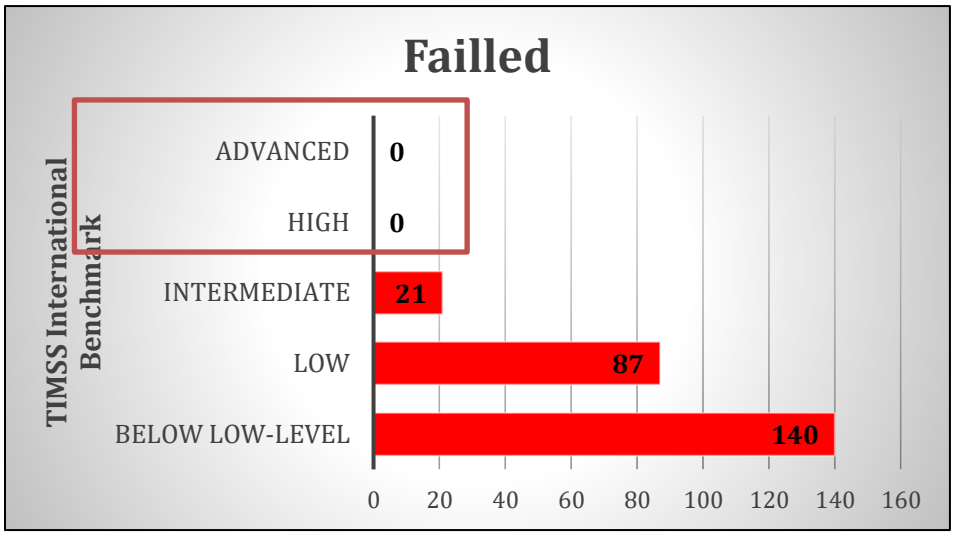

Figure 4. Distribution of the unsuccessful group according to TIMSS international benchmarks

Figure 4 shows that 56.5% (N=140) of the 248 students in the unsuccessful group were below the Low Level, 35.1% (N=87) were at the Low Level, and 8.5% (N=21) were at the Intermediate TIMSS international proficiency level, and there were no students at the Upper and Advanced Levels.

The contingency table from a pass/fail classification of students based on the low level of TIMSS international proficiency levels (400) and a pass/fail classification based on the cut-off score (45.49) determined by ANN for the Booklet-11 mathematics subtest is given in Table 2.

Table 2. Mathematics subtest (Booklet-11) contingency table

		TIMSS International Benchmark (400)		
		Passed	**Failed**	**Total**
Math. Test (45.49)	**Passed**	191	2	**193**
	Failed	108	140	**248**
	Total	**299**	**142**	**411**

As presented in Table 2, according to the pass/fail classification using the lowest TIMSS 400 International Benchmark criterion and the criterion of 45.49 for the Mathematics subtest, the number of people who passed the mathematics test in both criteria was 191 and the number of people who failed was 140. Using the value in Table 2, Sensitivity (true positive rate), Specificity (true positive rate), and Accuracy values can be obtained as the agreement values of the two criteria. These values are presented in Table 3.

Table 3. Math subtest (Booklet-11) fit values

Agreement for the passed	**Sensitivity = 191/299 = 0.64**
Agreement for the failed	Specificity = 140/142 = 0.99
Overall agreement	Accuracy = (140 + 191)/411 = 331/411 = 0.81

The agreement for the success case was 64% according to both criteria, while the agreement for the failure case was 99%. The overall agreement level of the pass/fail classification using both criteria was 81%.

Then, within the scope of the study, the answer to the question "What is the most appropriate cut-off score to be used for the evaluation of the scores obtained from the SLLMS in the TIMSS 2015 8th grade student questionnaire by ANN analysis method?" was sought. Figure 5 shows the cut-off score determined by the ANN analysis method for the decision of likes/dislikes learning mathematics to be made according to the total score of the students from the SLLMS, which consists of 9 items and is graded between 1-4.

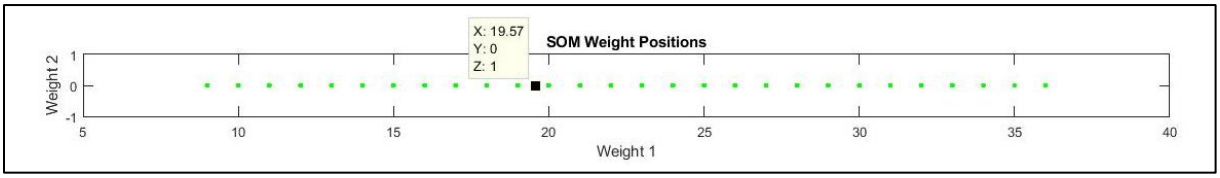

Figure 5. Cut-off score determined for SLLMS scale total score

Figure 5 indicates that the most appropriate cut-off score to be used for the evaluation of the total scores obtained from the SLLMS in the TIMSS 2015 8th grade level student questionnaire with the ANN analysis method was 19.57 over 9-36 points. According to the determined cut-off score, the SLLMS status of 5,741 students in the study group is shown in Figure 6.

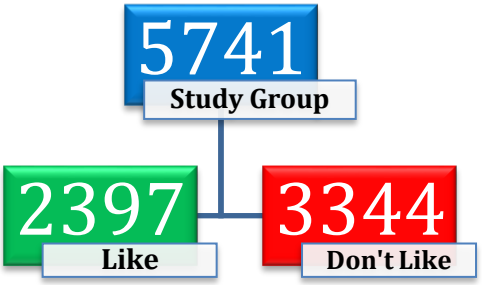

Figure 6. The status of the students in the study group
for liking learning mathematics according to the SLLMS

According to the cut-off score determined, 2,397 students (42%) out of 5,741 students in the study liked learning mathematics, while 3,344 students (58%) did not like learning mathematics.

Within the scope of the study's second aim, the answer to the question "How is the agreement between the SLLMS cut-off score determined by TIMSS guidelines and the SLLMS cut-off score determined by ANN analysis method?" was sought. Table 4 shows the contingency table for the classification of liking/disliking learning mathematics according to the 22-point criteria determined based on the statement given in the original description of the SLLMS scale and the 19.57-point criteria determined by the ANN analysis method.

Table 4. SLLMS contingency table

		TIMSS Description Cut Score (22)		
		Like	**Don't Like**	**Total**
SLLMS ANN (19.57)	**Like**	1440	957	**2397**
	Don't Like	0	3344	**3344**
	Total	**1440**	**4301**	**5741**

Table 4 reveals that the number of people who liked learning mathematics in both criteria was 1,440, and the number of people who did not like learning mathematics was 3,344 according to the classification of liking/disliking learning mathematics using the 22 score criteria based on the TIMSS description and the 19.57 criterion obtained from ANN analysis. Table 5 shows the agreement values calculated with the values in the contingency table.

Table 5. SLLMS agreement values

Agreement for liking to learn mathematics	**Sensitivity = 1440/1440 = 1.00**
Agreement for not liking learning mathematics	Specificity = 3344/4301 = 0.78
Overall agreement	Accuracy = (1440 + 3344)/5741 = 4784/5741 = 0.83

When Table 5 is examined, it is seen that the level of agreement in the case of liking to learn mathematics according to both criteria was 100%, while the level of agreement in the case of disliking to learn mathematics was 78%. The overall level of agreement for the classification of liking/disliking learning mathematics using both criteria was 83%.

According to all the findings obtained, it can be said that the validity of the cut-off scores determined by ANN analysis for the Mathematics subtest Booklet-11 and for the SLLMS scale was high.

3. Discussion, conclusion, and recommendations

The current study basically aimed to show how to determine the cut-off score to be used in the interpretation of the total scores obtained from an achievement test or scale with the Artificial Neural Network method. In addition, the validity of the cut-off scores determined within the scope of the study was also examined.

As a result of the study, the most appropriate cut-off score to be used for the evaluation of the total scores obtained from the TIMSS 2015 8th grade level Booklet-11 Mathematics subtest was determined as 45.49 out of 0-100 points with the Artificial Neural Network analysis method. According to the cut-off score, it was concluded that 193 (47%) of the 411 students in the study were successful, and 248 (53%) were unsuccessful. According to the lowest TIMSS International Benchmark (TIMSS International Benchmark) of 400 points and the 45.49 criterion determined by ANN for the Booklet-11 mathematics subtest, it was concluded that the agreement was 64% in the case of success and 99% in the case of failure. The overall agreement level of the pass/fail classification using both criteria was 81%.

With the artificial neural network analysis method, the most appropriate cut-off score to be used for the evaluation of the scores obtained from the SLLMS in the TIMSS 2015 8th grade student survey was determined as 19.57 over 9-36 points. According to the cut-off score, 41.8% (N=2,397) of the students liked learning mathematics, while 58.2% (N=3,344) did not like learning mathematics. According to the 22-point criteria determined on the basis of the expression given in the original description of the SLLMS and the 19.57-point criteria determined by the ANN analysis method, it was concluded that the agreement was 100% for the case of liking to learn mathematics and 78% for the case of disliking to learn mathematics. The overall level of agreement for the classification of liking/disliking learning mathematics using both criteria was found to be 83%. Birican et al. (2010) state that SOM-type networks are ideal for cluster analysis; however, it may be necessary to consult expert opinion on the subject to check the accuracy of the results

obtained. The study not only determined the cut-off score, but also examined the validity of the cut-off score, which eliminated the need to apply for expert opinion on the results obtained.

The study results suggested that the validity of the standard-setting studies conducted with the Artificial Neural Networks method was high. Therefore, researchers are recommended to use the Artificial Neural Networks method in determining the cut-off score to be used in interpreting the total scores obtained from the achievement test or the total scale scores obtained from the scales. Because traditional standard-setting methods involve expert opinion, subjectivity may be in question. However, there is no subjectivity since the ANN method does not require expert opinion. In the study, a cut-off score was determined using the ANN method. In future studies, cut-off scores can be determined using different standard-setting methods, and the cut-off scores' validity can be examined.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public commercial, or not-for-profit sectors.

The author declares no competing interests.

References

Bircan, H., Zontul, M., & Yüksel, A. G. (2010). SOM tipinde yapay sinir ağlarını kullanarak Türkiye'nin ihracat yaptığı ülkelerin kümelenmesi üzerine bir çalışma [A study on the clustering of countries to which Turkey exports using SOM type artificial neural networks.]. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, *20*(2), 219-239.

Gorr, W. L., Nagin, D., & Szczypula, J. (1994). Comparative study of artificial neural network and statistical models for predicting student grade point averages. *International Journal of Forecasting*, *10*(1), 17-34.

Hardgrave, B. C., Wilson, R. L., & Walstrom, K. A. (1994). Predicting graduate student success: A comparison of neural networks and traditional techniques. *Computers & Operations Research*, *21*(3), 249-263.

Haykin, S. (1999). *Neural networks: A comprehensive foundation*. New Jersey: Prentice Hall International.

Ibrahim, Z., & Rusli, D. (2007, September). Predicting students' academic performance: comparing artificial neural network, decision tree and linear regression. In *21st Annual SAS Malaysia Forum* (pp. 1-6).

Kohonen, T. (2001). *Self-organizing maps*. Berlin: Springer.

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). *TIMSS 2015 International Results in Mathematics*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2015/international-results/.

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/international-results/.

Öztemel, E. (2003). *Yapay sinir ağları* [Artificial neural networks]. İstanbul: Papatya Yayıncılık.

Saraç, T. (2005). *Estimating software project duration using artificial neural networks* [Unpublished doctoral dissertation]. Gazi University.

Shah, J. R., & Murtaza, M. B. (2000). A neural network based clustering procedure for bankruptcy prediction. *American Business Review*, *18*(2).

SubbaNarasimha, P. N., Arinze, B., & Anandarajan, M. (2000). The predictive accuracy of artificial neural networks and multiple regression in the case of skewed data: Exploration of some issues. *Expert systems with Applications*, *19*(2), 117-123.