Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
42(4), Fall 2023, pp. 1-21
https://doi.org/10.22099/tesl.2023.47604.3194

Online ISSN: 2717-1604
Print ISSN: 2008-8191
Research Paper

Intelligent Scoring in an English Reading Comprehension Course Using Artificial Neural Networks and Neuro-fuzzy Systems

Mohamad Reza Farangi *　　　　**Iman Zabbah** **

Abstract

Predicting students' performance in a course is one of the major aims of educational data mining systems. In the present study, two and three-layer artificial neural networks (ANN) and neuro-fuzzy systems (NFS) were used to predict Iranian EFL learners' final scores and compare them with scores given by their instructor. Sixty-six students' scores in an English reading comprehension course comprising of five sub-scores of midterm (out of 40), quiz (out of 60), final (out of 50), class participation (out of 5) and bonus (out of 2) were used for training the systems. Two and three-layer ANNs and an NFS were trained to predict students' final scores using training data. Researchers compared the students' final scores given by their instructor and those achieved through the ANNs and NFS. The results showed that the NFS could predict and deliver scores that were closer to the linear sum of students' scores. Moreover, three-layer ANN had a better performance than the two-layer ANN. According to these results, data mining techniques could deliver an accurate estimate of students' abilities in a particular course.

Keywords: Educational Data Mining, ANN, NFS, Subjective Scoring, Intelligent Scoring, Reading Comprehension

Considering the important role of the increased speed of computers in the advancement of data mining science, it can be said data mining in educational systems can encompass a wide range of fields (Bagheri Nevisi & Arab, 2023). Generally, in recent years, the use of data mining for planning and education purposes has gained a lot of importance (Adekitan & Salau, 2019; Bousbia & Belamri, 2014). Three main purposes purported for educational data mining (EDM) are predicting students' performance based on different algorithms, defining the elements used in predictions, and quantifying aspects of students' performance at the course or program levels (Mehdi & Nachouki, 2022). Data

* Received: 10/06/2023　　　　Accepted: 26/11/2023

* PhD, Department of Language and Linguistics, Shiraz University, Shiraz, Iran, mohamadrezafarangi@gmail.com (Corresponding author)
** Assistant professor, Faculty of Computer engineering, Islamic Azad university, Torbati-e-Heydariyeh Branch

How to cite this article:

Farangi, M., & Zabbah, I. (2023). Intelligent Scoring in an English Reading Comprehension Course Using Artificial Neural Networks and Neuro-fuzzy Systems. *Teaching English as a Second Language Quarterly (Formerly Journal of Teaching Language Skills)*, 42(4), 1-21. doi: 10.22099/tesl.2023.47604.3194

mining techniques have enabled academics to analyze educational data and extract useful information, which can assist academic advisors in comprehending students' learning processes and advancing their knowledge and skills (Mehdi & Nachouki, 2022). Adekitan and Salau (2019) used the Konstanz Information Miner (KNIME) based data mining model to predict the fifth year and final GPS of Nigerian University students. The results showed an 89.15% accuracy of six predicting models, so the researchers suggested that these models can be used to identify potential future poor results or not graduating at all. Sultana et al. (2019) used EDM to predict students' performance and learning behavior by uncovering hidden knowledge from learning records or educational data. Using various classification systems, they investigated student performance and found that Deep Neural Networks and Decision Trees produced the best results. They argued that the huge bulks of educational data may make it frustrating for a single teacher to predict or assess students' performance so teachers can take advantage of EDM services.

Similarly, scoring systems can take great advantage of data mining capabilities (Haghi et al., 2023). The traditional scoring system is based on a subjective sum of several sub-scores. It can be subjective depending on several variables, including gender, race, ethnicity, political issues, etc. Comparing traditional scoring systems (subjective scoring) with the use of artificial systems through data mining (objective scoring) can provide teachers, scholars and educational policymakers with insightful ideas in this regard (Tomasevic et al., 2020). Liu et al. (2014) meta-analysis study showed a moderate to almost complete agreement between computer and human scorings, while Shermis (2015) could not find such an agreement. Bridgeman et al. (2012) showed that demographic characteristics such as language, gender and ethnicity can lead to variances between computer and human scoring. A few studies on automated scoring of assessments showed a positive correlation between computer and human scoring in the domains of mathematics and sciences (Kersting et al., 2014; Wilson et al., 2017).

As a matter of fact, the present study compared a traditional scoring system with artificial intelligence systems scoring. In the traditional scoring system, different subjective weights are given to each sub-score. For example, the final score in a course is comprised of a sum of quizzes, midterms, final exams, projects, etc. So, the teacher may give a higher weight to midterm for one student and to quizzes for another student. In this way, the final score is a linear sum of these sub-scores plus subjective variables, as mentioned above. However, an intelligent system can be trained to give different weights to each sub-score and then estimate the final score based only on those weights, not including the subjective variables. This way, it is believed that a fairer evaluation of students' performance can be achieved.

Conceptual framework

Exploring hidden data in educational systems can help decision-makers in the field of education to improve and enhance educational processes such as planning, registration, evaluation and consulting. Considering the fact that the educational system is always faced with a lot of data and information about educational centers, students, professors, personnel, etc., and paying attention to their educational performance is an integral part of this system, educational data are really valuable; As a result, the use of data mining in the educational system can have many applications and advantages. Over the past decades, data mining techniques have been able to solve real-world problems successfully (Talpur et al., 2023). Artificial neural networks (ANNs), Fuzzy logic, swarm intelligence, genetic programming, and hybrid approaches such as neuro-fuzzy and genetic fuzzy systems are examples of data mining programs all leading to the design and analysis of complex, intelligent systems (Talpur et al., 2023). To take advantage of huge learner-related data generated during learning and education, artificial neural networks, as data-based methods, can approach function approximation (regression), pattern recognition (classification), and predictive analytics to forecast learners' performance and classify their behaviors with the purpose of improving retention, progression, motivation and cost saving (Brocardo et al., 2017). In addition, one of the primary purposes of all these approaches is to introduce objectivity (i.e., staying away from biases) in the realm of education. However, the question is, "Does objectivity really exist?" The argument in this paper relates to the fact that teachers' subjective decisions in the process and product of assessment may endanger assessment reliability and validity. Meanwhile, as argued by Virk et al. (2020), assessment cannot be purely objective. Every assessment practice is led by the beliefs, opinions, values and conceptions of the assessor(s) and expecting a purely objective process is a mystery. What is actually done is to make subjective decisions based on the supposed objective processes. This process is called objectification by Vlueten (Van der Vleuten, 1999). In spite of that, academia should strive for an assessment process that is objective to the extent possible. This can be achieved through the use of artificial intelligence systems, which have been trained to examine assessment results and decisions based on predetermined criteria.

Review of the Literature

The literature regarding artificial intelligence systems is varied. The main focus in the literature is delivering systems with high prediction capabilities. These systems are generally used as supplements to human-made decisions. The studies which are reviewed here focused on the educational uses of ANN and NFS systems.

Saa et al. (2020) used DTs, naive Bayes, and artificial neural network algorithms to predict student academic performance from a new dataset at a private university in the United Arab Emirates (UAE). The Random Forest (RF) algorithm used in this study could predict students' performance carefully. The researchers suggested that the ANN system can play an important role in identifying students' weaknesses and the factors affecting them. Similarly, Ahajjam et al. (2022) predicted students' final scores using various Machine Learning (ML) algorithms. The data included students' grades at the common core, first-year baccalaureate, and second-year baccalaureate levels. These data were used to predict the baccalaureate average. The results provided the researchers with acceptable scores and predictions. Both studies suggested the efficient functioning of ANNs in predicting students' academic performance.

Besides academic achievement, other variables were also predicted successfully. Iatrellis et al. (2021) presented two methods of unsupervised (K-means algorithm) and supervised (random forest algorithm) learning techniques to predict time to degree completion and students' enrollment in a computer science curriculum in Greece. The results indicated predictions with relatively high accuracy. In a similar attempt, Fernandes et al. (2019) presented a descriptive and preventive statistical study using data mining algorithms to predict the performance of Bronze Capital students via classification by the Gradient Boosting Machine (GBM) algorithm, which showed that the grades and absences attributes were the most relevant factors in predicting the end-of-year academic results of student performance.

Further, Tomasevic et al. (2020) compared machine learning algorithms to predict student performance in exams, i.e., finding students with a "high risk" of dropping out of the course. Students' engagement data and their past performances were fed to ANNs, and the outcome was a high precision in the predictions. In the same vein, Uskov et al. (2019) proposed a project that evaluated eight ML algorithms to predict students' academic performance in STEM courses. The researchers argued that the ANNs could produce accurate predictions and that numeric values (from 0 to 20) were more accurate systems of grading than the letter grades (A, A+, B, etc.) in these systems. On the whole, these studies approved ANN applications in educational predictions. However, due to their limitations, neural networks have been complemented by fuzzy systems.

With the emergence of deep learning concepts, some researchers introduced fuzzy inference system elements and modules into such systems to address possible uncertainties in raw data. According to Göktepe Yıldız and Göktepe Körpeoğlu (2023), fuzzy systems lack learning abilities. By combining fuzzy rules and neural networks, the evaluation tool may have greater adaptability to changing conditions. Besides, neuro-fuzzy systems (NFS) help specialists to produce rule-based structures (Talpur et al., 2023). These systems use rule-based structures to help an analyst easily understand how a decision has been made (An et al., 2019). They can be applied to various constructs involving encoding both objective measurements and subjective information (Bonanno et al. 2017). Basic neuro-fuzzy systems have been major research topics for over three decades, and various surveys and systematic reviews can be found in the literature. For example, Son and Fujita (2019) attempted to predict university students' future performance through the MANFIS-S method through local and global training. The results showed that this method was successful in predicting students' future performance and outperformed several neural network algorithms. Likewise, Göktepe Yıldız and Göktepe Körpeoğlu (2023) predicted problem-solving perceptions of 360 students in Turkey through neuro-fuzzy and hierarchical regression, and the results showed that the Adaptive neuro-fuzzy inference system (ANFIS) could accurately predict students' scores. As a surprising result, the researchers argued that the environment played a role in these predictions. In addition, Mehdi and Nachouki (2022) predicted students' grade point average (GPA) using ANFIS at Ajman University, Oman. The results showed that the ANFIS method had a better prediction than the conventional multilinear regression. These and some other studies showed better performances of NFS over ANN systems. Therefore, the present study followed the same line of research to compare them.

Objectives and Research Question

In order to increase educational standards, we need data mining systems that provide the knowledge and insights needed for decision-makers in the educational system. Academic success of students is one of the most important things in the field of education so it is necessary to use hidden relationships between data bundles to predict students' success factors. Unfortunately, despite the abundance of data available in the education system of Iranian universities, a deep and comprehensive investigation has never been done to extract hidden knowledge from these data in the field of applied linguistics. We aimed to compensate for this knowledge gap in the field of Applied Linguistics. Hidden patterns, dependencies and exceptions discovered by some data mining techniques can be used to improve the efficiency, effectiveness and speed of processes. As a result, this improvement has many advantages, such as maximizing the efficiency of the educational system, reducing the rate of loss and exclusion of students, increasing the rate of students' development, reducing students' retention duration, increasing students' success, increasing student learning output and reducing the cost of system processes for the education system (Adekitan & Salau, 2019). As a matter of fact, the knowledge that can be discovered through data mining in the field of education can be used not only by the owners of the system, i.e., teachers and educational officials but also by the users of the system, i.e., students (Ranjan & Malik, 2007).

Data mining has been recently practiced in engineering (Adekitan et al., 2019), marketing and product design (Jin et al., 2019), business management (Zuo et al., 2016), computer science (Mahendra et al., 2019), biological studies (Gu et al., 2018), genetics (Noreña et al., 2018), health and drug development studies (Keserci et al., 2017), facility maintenance management (Miguel-Cruz et al., 2019), meteorology (Kovalchuk et al., 2019), chemistry and toxicity analysis (Saini & Srivastava, 2019), transportation safety (Divya et al., 2019) and traffic management (Amiruzzaman, 2019). In the field of education, data mining has been used to find trends related to student performance (Adekitan & Noma-Osaghae, 2018; Roy & Garg, 2018), learning behaviors (Kim et al., 2018), and student potential prediction (Yang & Li, 2018; Ibrahim et al., 2019). Meanwhile, there is a gap in using data mining techniques in language testing which may have wider consequences in students' lives. In this regard, the objectives of this study can be described as below:

The present study used data mining techniques to predict students' final scores and compare them with teacher's scores. For this purpose, the researchers collected and processed educational data using data mining techniques for the purpose of improvement in language testing procedures with the aim of reaching an optimal data mining model based on evaluation factors for the prediction of students' scores. As a result, this study is one of the first few studies in applied linguistics in which the researchers modeled and trained intelligent systems to take into account subjective parameters in teachers' scoring and predict students' scores. For this purpose, the following research questions are proposed:

RQ1. Can data mining techniques (NFS and ANN systems) predict teacher's scoring behaviors?

RQ2. Which technique has a higher accuracy in predicting students' scores?

Method

Data preparation stage

The process of data mining in the systems includes (1) data collection, (2) data pre-processing, (3) application of data mining algorithms, (4) and post-processing (Garcia et al., 2007). The data in this study were collected from the score checklist of one of the Iranian university professors. The course was reading comprehension in an English as a Foreign Language department. Sixty-six students' scores comprising five sub-sections were collected in the form of an Excel sheet (see Appendix). These data were related to male and female students whose ages ranged from 22 to 29 years old (m= 23.8). They studied English language teaching as a field of study. Their instructor held a Ph.D. in Applied linguistics. He had 7 years of experience teaching BA, MA and PhD students in English language teaching and English Literature. He was a full-time faculty member at Shiraz University, Shiraz, Iran. The students' final scores consisted of 5 sub-sections, including midterm (out of 40), quiz (out of 60), final (out of 50), class participation (out of 5) and bonus (out of 2). In his mind, the teacher may have given different weights to these sub-scores in order to obtain the final score, so it was not a direct sum of sub-scores. Considering that subjectivity can affect educational decisions, linear modeling may not be efficient in achieving reliable and valid final grades, so intelligent modeling is used. One of the best intelligent modeling methods is an artificial neural network, which is described below. The ethical approval for this study was collected through the consent form. For this purpose, the students in the reading comprehension course agreed to share their scores with the researchers in this study with the condition of anonymity, and the instructor of the course also filled out the consent form with the condition of confidentiality of students' information.

Network Designs

Artificial Neural Network (ANN), an artificial intelligence modeling technique, is generally used to work on approximation and pattern classification problems (Khan et al. 2021). Some ANN models are proposed in the literature, including Radial Basis Function models, Multilayer Perception Models, Generalized Neural Network models, etc. Among these models, the MLP ANN models have stood out most recently because they are very robust and popular for learning, function approximation, and pattern classification (Guo et al., 2014; Jo et al., 2020). The MLP possesses many robust algorithms that can be explored to carry out more proficient adaptive non-subjective statistical modeling over the classical logistic regression methods that are frequently engaged in developing predictive models (Tataria et al. 2021).

MLP consists of an input layer, one or several hidden layers and an output layer (Isabona et al., 2022). In this structure, all neurons of one layer are connected to all neurons of the next layer. This arrangement is a network with full connectivity. The network uses two types of different signals. The first are the signals calculated based on the inputs of each neuron, weight parameters and its stimulus function, and the second are error signals calculated by returning from the output layer and dividing into the other hidden layers (Isabona et al. 2022). The number of hidden layer neurons depends on the network designer's view and is obtained with trial and error. In the absence of sufficient neurons, the network will not be able to make accurate mapping between input and output vectors (Isabona et al., 2022). At the output of each neuron in the MLP network, there is a subjective function, and the learning process takes place in all neurons and layers. All the weights and biases of the network can change during the learning process (Isabona et al., 2022). Figure 1 shows the 3-layer neural network structure.

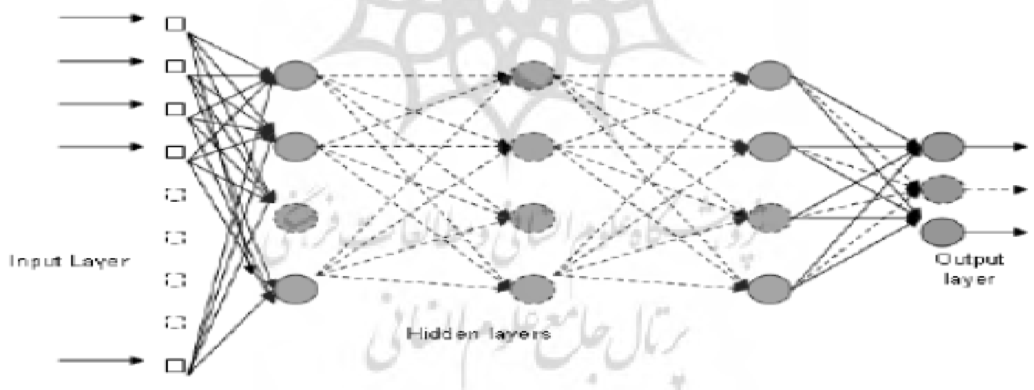


Figure 1. *Multilayer perceptron neural network architecture used in this study*
Source: Isabona et al. (2022)

Neurons of different layers are connected by a series of weighted connections. The first layer is the input layer, where each neuron is associated with an input variable. The last layer is the output layer, which is attributed to the predictor variable or variables. In between, there are a number of intermediate or hidden layers. The input nodes are connected to a number of nodes in the hidden layer. Nodes in the hidden layer can be connected to nodes in another hidden layer or to the output layer. The number of input nodes depends on the number and type of features of the data set, and the number of output nodes depends on the type of classification operation. However, the number of nodes in the hidden layer is an innovative concept and is obtained by trial and error.

In the present study, all variables were set in the range of (-1, +1) through normalization. By giving the initial random weights and mixing the records randomly, we were able to increase the accuracy of the model. The training time of the model was 500 rounds, and the learning rate was 0.3. However, to prevent the construction of a network with a high error rate, an evaluation set of 25% of the training data was used. The training of the model continued until either the error rate of the evaluation set deteriorated continuously or the training time ended. To build the optimal neural network model, we started with zero hidden layers and gradually increased the number of hidden layers. In fact, by increasing the number of layers, the nonlinearity and complexity of the model increase, and this increase in complexity ends up at the cost of overfitting and reducing the accuracy of the model in the evaluation set. In the case of our data, with the increased number of layers, the accuracy of the model increased regularly. The number of hidden layers in the final model was 3 layers. To classify the data using neural networks, the backpropagation algorithm was used, and the activation function of all nodes was sigmoid. Five features have been used to build networks because removing the features would reduce the accuracy of the network. It should be noted that to build neural network models, it is necessary to convert all nominal variables into numerical values.
Equation 1 shows the output of the last layer.

$$O_i = sgm\left(\sum_m sgm\left(\sum_l x_i w_{lm}^h\right) w_{mi}^o\right) \quad (1)$$

The output of the network in the last layer was calculated by formula 1, where *h* and *o* represent the hidden layer and the output layer, respectively, and *w* represents the weights of the layers.

This network used general approximation capability and Feed Forward backpropagation learning algorithm. Also, the gradient descent method of tnecseD neidarG was used to minimize the squared error between the output of the network and the desired response. To update the weights due to fast convergence, the Lunberg-Marquardt algorithm has been used.

Neuro-fuzzy system design

First, a two-layer network (one hidden layer and one output layer) was designed. Mid, quiz, final, class participation and bonus parameters were applied as input and the total score parameter as output to the neural network. Usually, in all neural networks, 80% of the data, i.e., 56 samples, were used as train data and 20% of them, equivalent to 14 samples, were used as test data. Since the score estimation is highly dependent on the artificial neural network architecture, about 50 neural network architectures were examined, and the error of each one was calculated and shown in Figure 2.

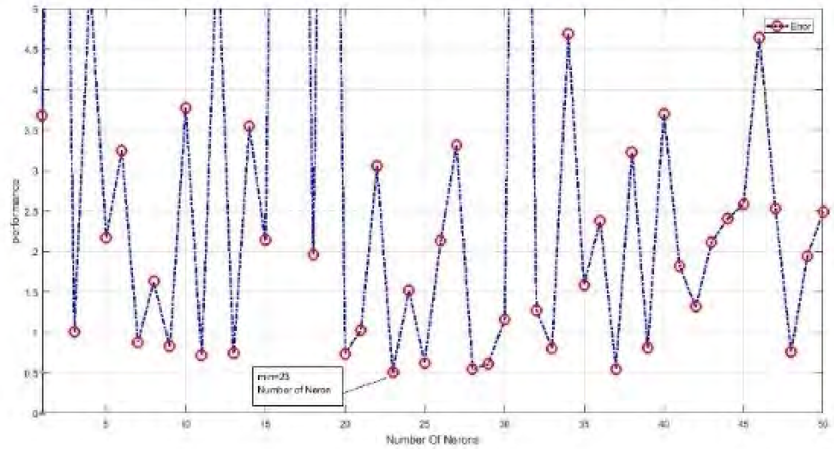


Figure 2. *Number of neurons*

As Figure 2 shows, the number of neurons in the first hidden layer varied between 1 and 50, and the minimum error occurred when the number of neurons in the middle layer was 23. To ensure the correctness of the output system of the artificial neural network, it has been compared with the real output, i.e. the grades assigned by the instructor to the student. The mean error obtained in this method using the least square error method was 0.5. This means that the designed system was able to instructor’s grading system and was able to estimate the final score for 10 students who were not in the training samples with a difference of 0.5.

Next, a three-layer network (two hidden layers and one output layer) was designed. Since the difference between the student's actual score and the score predicted by the system could be improved, we changed the structure of the neural network and increased the number of hidden layers. The results were analyzed for 10 multilayer perceptron neural networks. The error obtained in this model has decreased compared to the previous method and has reached 0.03. The results of different neural network architectures and the lowest error related to each network were recorded in Table 1.

Table 1.
Examination of different neural network architectures considering continuous parameters

RMSE among 66 individuals	Number of hidden layers	Neural network architecture
0.2882	2	-1j5-10-
0.2866	2	-1j5-20-
0.2795	2	-1j5-30-
0.2574	2	-1j5-40-

The fold-k cross-validation method was used to validate the results. If we randomly separate the training data set into subsamples with the same size at each stage of the validation process, a number of these layers can be considered as the training data set and one as the validation data set. In this research, we considered K=10.

Neuro-fuzzy system design

In the next stage of the study, a fuzzy system was used in conditions of uncertainty because fuzzy rules are suitable for such conditions. Since there is uncertainty in the nature of the problem in assigning a student's grade, a fuzzy view of the problem can be a qualitative approach in the evaluation of students, which is preferable to a quantitative system. Neural and fuzzy system combination leads to the production of fuzzy-neural systems that will be closer to the human reasoning style. The neural fuzzy classification algorithm used in this research was one of the fuzzy system adjustment methods based on which rules can be added and removed. This system was a special five-layer forward propagation network where the first layer represented the input variables. The second and third layers represent the hidden layer, which includes fuzzy rules and uses *T_norm* as the activation function. The fourth layer showed the output layer, and finally the fifth layer also showed the output variable. Figure 3 shows the designed fuzzy neural network.

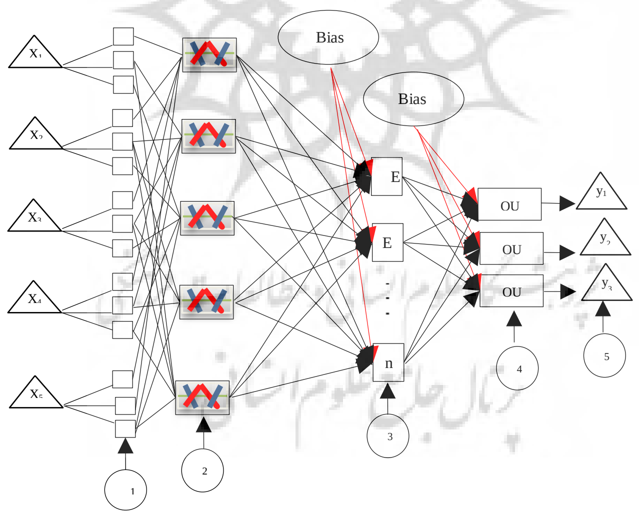


Figure 3. *A neuro-fuzzy network designed for fuzzy assessment of students' scores*

In the current study, five continuous input variables, including mid, quiz, final, class participation and bonus and an output variable were selected for network design, and the linguistics variable required for each variable has been created with the help of an expert (professor) as shown in Table 3.

Table 3.
Linguistic variables assigned to grades

	Range	Very weak	Weak	Good	Very Good	Excellent
Midterm	0-40	0-8	9-18	19-28	29-35	35-40
Quiz	0-55	0-10	11-25	26-35	36-45	46-55
Final	0-50	0-10	11-25	26-35	36-45	45-50
Class Participation	0-5	0-1	1-2	2-3	3-4	4-5
Bonus	0-2	0-0.4	0.5-0.9	1-1.4	1.5-1.7	1.8-2
Total	0-20	0-10	10-12	13-15	16-18	19-20

The process of fuzzification of all inputs and outputs and determining the membership function was done in the next stage of the study. For the defuzzification process, the designed system was based on the central method. Each input variable has its own number of membership functions that use linguistic variables for the fuzzification strategy. The set of fuzzy rules is determined after determining the input and output membership functions. Figure 4 shows the set of input variables to the fuzzy neural system.

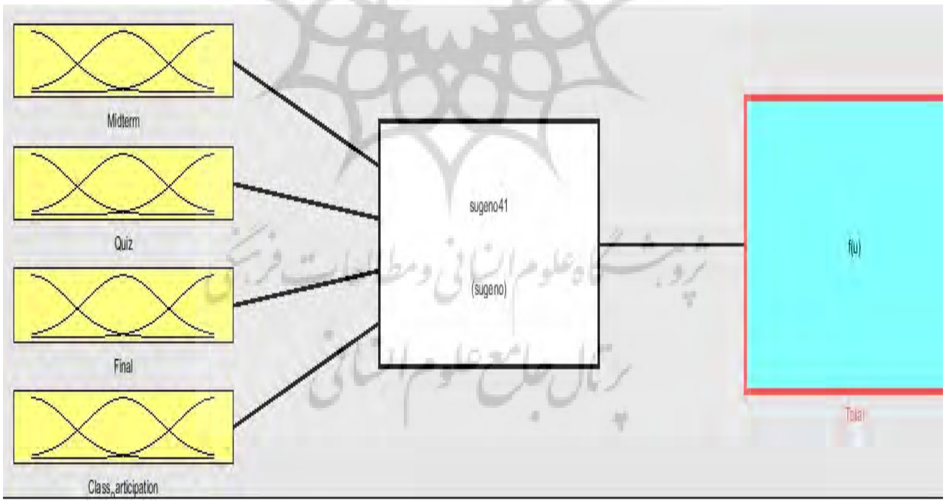
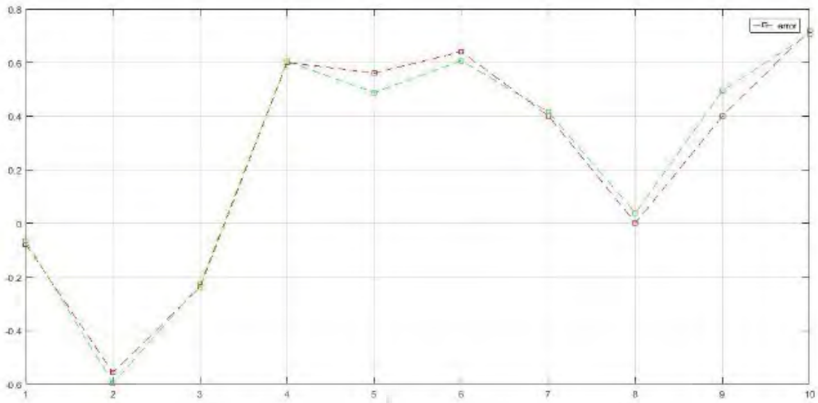


Figure 4. *Input variables to the Fuzzy system*

The Sogno algorithm was used in the design of the neuro-fuzzy network. According to input variables, the neuro-fuzzy network can have a different number of membership functions. In the current research, about 70 different types of membership functions were examined, and in the best case, with the repetitions of 100, the average error reached 0.0019. Figure 5 shows the real scores and the predicted scores of students using the fuzzy neural method.

Figure 5.
Real and predicted scores



Also, Figure 6 shows the actual and predicted grades of 66 students in the class with the NFS model.

Figure 6.
Real and predicted grades for the class

In the above figure, students' real scores are shown in blue, and their NFS-predicted scores by red.

Results and Discussion

In this research, the error rate criterion obtained from the Fold-K validation method and the error obtained from testing the models on the evaluation set were used to compare the prediction results. The information related to the level of accuracy of the different models built in 3 modes of the evaluation set is given in the following table.

Table 4.
The error rate of different models

	Two-layer model error rate	Three-layer model error rate	Neuro-fuzzy model error rate
S1	0/14689	0/136592	0/082634
S2	0/290834	0/152823	0/061118
S3	0/166751	0/16009	0/084716
S4	0/165604	0/121106	0/121062
S5	0/518003	0/120148	0/042735
S6	0/446624	0/278696	0/052085
S7	0/394926	0/139471	0/044488
S8	0/183658	0/112376	0/158118
S9	0/125759	0/102643	0/078771
S10	0/39045	0/147271	0/043871

In order to picturize data, the following figure provided information in the form of columns.

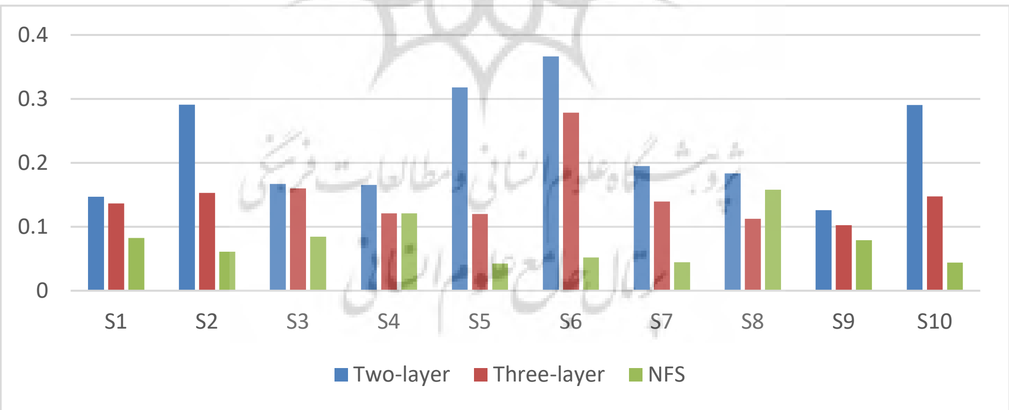


Figure 9. *The error rate of different models*

As shown in the above chart, the error rate of the NFS model was lower than the other models for all 10 samples in this case. The two-layer model had the highest error rate, and the three-layer model was in the middle. This can be explained in light of the fact that NFS is a more complicated model, taking advantage of the learning power in the ANNs and the functionality of fuzzy systems. In this way, it is believed that the reasoning and inference capabilities of the system increase and therefore, it can deliver more accurate results. As shown in Table 4, the accuracy of the three models was examined through the root mean square error (RMSE) and based on these values, the NFS model had a higher accuracy than the other models. For S1 and S3, the two-layer and three-layer models had quite similar error rates. For S4, the error rate of the three-layer ANN and NFS were similar and more importantly, the error rate of NFS was higher than the three-layer ANN for S8. S6 showed the highest prediction error in both two-layer and three-layer ANN models. For S2, the two-layer ANN performed very low in predicting students' scores, while the three-layer and NFS models performed closer to the real scores. The teacher's scores and the scores predicted by the models were compared in the following table.

Table 5.
Teacher scores compared with the prediction model scores

	Prediction accuracy	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**	**S10**
Teacher Scores	0	13	10	12	17.25	17	17.5	16	13.5	16	18
two-layer neural network model	79.6%	13.98	9.26	12.84	16.34	16.47	16.90	15.58	13.99	16.44	16.99
Three-layer neural network model	82.3%	13.63	9.67	12.82	16.63	16.76	17.11	15.66	13.77	16.23	17.31
neuro-fuzzy network model	87.8%	13.44	9.81	12.53	16.82	17.02	17.33	15.88	13.82	16.66	17.73

In Table 5, a sample of the scores given by the teacher were compared with the scores given by the two-layer, three-layer and neuro-fuzzy models, respectively. The majority of scores predicted by the systems were higher than the teacher scores. This can be considered an interesting finding as the researchers can argue that the teacher has underestimated his/her students' performance in some way which was not directly related to their scores. However, for one student (S2) who was at the edge of failing the course, the teacher overestimated his/her score (maybe to help the student), but the systems predicted scores that were lower than 10.

Generally, it can be observed that the models have been able to learn and predict teacher's scoring behavior. The neuro-fuzzy model had the most similar performance with the teacher's scoring behavior. Using the NFS model to predict students' scores showed satisfactory results. The average error rate in 3-layer validation and testing on the evaluation set was lower than other methods. Since the main goal of prediction models was predicting students' scores, it is necessary to compare the models based on the correct classification of these samples. The results of this study showed that the ANN and NFS could predict students' reading comprehension scores. However, the NFS performed better than the two-layer and three-layer ANN in predicting students' scores. These results were in line with the results of Iatrellis et al. (2021), who trained machine learning models and showed that these models could accurately predict students' scores. Similarly, Göktepe Yıldız and Göktepe Körpeoğlu (2023) showed that the adaptive neuro-fuzzy inference system could accurately predict students' scores. In another similar study, Mehdi and Nachouki (2022) showed that the adaptive neuro-fuzzy inference system had a higher accuracy compared to the multilinear regression models in predicting university students' GPA. Son and Fujita (2019) concluded that the Multi Adaptive Neuro-Fuzzy Inference System with Representative Sets (MANFIS-S) algorithm was successful in predicting students' scores. In a hybrid method study, Yin et al. (2020) predicted new product sales, and the results showed that the hybrid model (clustering analysis + deep learning) could achieve higher accuracy in prediction.

In addition, these results showed that it is possible to model teachers' scoring behavior. As argued by Daniels and Bulut (2020), fair scoring is a process that may be threatened by several factors, one of the most important of which is teachers' subjective behavior. The authors believed that students sometimes express doubts about teachers' impartiality in scoring, and this may have sound consequences on their educational achievements. Actually, one of the main objectives of this study was to reduce or remove (if possible) subjectivity in scoring behavior. We wanted to portray a score for each student, which is indicative of his/her real and actual performance, not what happens in the teacher's mind. However, we did not and could not claim that these systems can replace teacher's scoring, but they can at least help teachers deliver a more reliable evaluation of their students.

As a matter of fact, the authors suggested computerized scoring as an acceptable assistant in this case. Finally, the reason for the advantage of NFS over the two-layer and three-layer ANNs may be that while the former uses a linguistics if-then approach, the latter is trained by data (Naresh et al., 2020). Accordingly, through a hybrid system, the researchers can use both if-then linguistics rules and real-world data, which, in turn, can enhance prediction accuracy.

Conclusion

In this article, an attempt was made to provide a suitable model for predicting students' final scores using students' records and with the help of different prediction techniques. The research question investigated whether or not NFS and ANN as artificial intelligence systems were able to predict students' scores in a reading comprehension course. Based on the results, the answer to this question was yes. Regarding the follow-up question, the results showed that the NFS was a more accurate predictor of the students' scores than the two- and three-layer ANNs.

Regarding the implications of the study, it can be argued that intelligent systems can facilitate our evaluation procedures and improve the accuracy and fairness of assessment processes. Prediction models can help lower the workload of teachers and ensure students that their academic progress will be judged without any bias. Moreover, Fuzzy modeling can help scholars approximate educational equality and justice, which may have profound educational advantages. The findings can also have implications for English language teachers. First, teachers should remember that their assessment decisions can be biased due to some subjective factors. Suppose they are working in contexts in which their decisions have high-stake consequences on students' lives. In that case, they should do their best to construct some personal or institutional assessment rubrics that are as objective as possible. Next, English teachers should get familiar with data mining systems and their applications in educational assessment. They can use these systems to deliver more objective assessment results to their students. In addition, English language faculties should consider some courses related to artificial intelligence systems in their curriculum. These courses can acquaint both instructors and students with the latest technological advances which can be used in language classrooms both for teaching and assessment purposes.

The researchers suggest that this line of research is enlightening for the future of assessment and scoring behavior. Moreover, the findings of the present study indicated the potential of data mining techniques in educational spheres. As discussed previously, the rich academic data (i.e., students' achievement scores, teacher evaluation data, students' GPA, etc.) in every field of study should not be ignored easily. These data have the potential to open new paths in educational assessment and evaluation, which a higher-order educational development and decision-making may follow.

One limitation is that this is a new domain, and this study only attempted to train the models and not to improve their performance. Therefore, future studies should strive to improve the prediction accuracy of models similar to this study. Another limitation is related to the scale of data. This was a small-scale study including data from around sixty-six students. However, EDM techniques have the potential to analyze huge data sets. Future researchers are recommended to replicate this study on larger data sets. Another suggestion for future studies relates to the combination of ANN and NFS systems, such as the ANFIS modeled by Daneshvar et al. (2021) for educational applications. Finally, it should be pointed out that pure objectivity is not possible (Tofighi & Ahmadi Safa, 2023) and even with the use of artificial systems, some hints of personal decisions can be tracked.

References

Adekitan A.I., & Salau, O. (2019). The impact of engineering students' performance in the first three years on their graduation result using educational data mining. *Heliyon*, *5*(2), e01250. https://doi.org/10.1016/j.heliyon.2019.e01250

Adekitan, A. I., Adewale, A. A., & Olaitan, A. (2019). Determining the operational status of a Three Phase Induction Motor using a predictive data mining model. International *Journal of Power Electronics and Drive System*, *10*(1), 91-103. http://doi.org/10.11591/ijpeds.v10.i1.pp93-103

Ahajjam, T., Moutaib, M., & Aissa, H., et al. (2022). Predicting Students’ Final Performance Using Artificial Neural Networks. *Big Data Mining and Analytics*, *5*(4), 294-301. https://doi.org/10.26599/BDMA.2021.9020030

Amiruzzaman, M. (2019). Prediction of traffic-violation using data mining techniques. In *Proceedings of the Future Technologies Conference (FTC) 2018: Volume 1* (pp. 283-297). Springer International Publishing.

An, J., Fu, L., Hu, M., Chen, W., & Zhan, J. (2019). A novel fuzzy-based convolutional neural network method for traffic flow prediction with uncertain traffic accident information. *IEEE Access, 7*, 20708–20722. https://doi.org/10.1109/ACCESS.2019.2896913

Bagheri Nevisi, R., & Arab, N. (2023). Computer-Generated vs. Direct Written Corrective Feedback and Iranian EFL Students’ Syntactic Accuracy and Complexity. *Teaching English as a Second Language Quarterly (Formerly Journal of Teaching Language Skills)*, *42*(2), 111-148. doi: 10.22099/tesl.2023.46955.3177

Bridgeman, B., Trapani, C., & Attali, Y. (2012). Comparison of human and machine scoring of essays: differences by gender, ethnicity, and country. *Applied Measurements in Education, 25*(1), 27-40. https://doi.org/10.1080/08957347.2012.635502

Bonanno, D., Nock, K., Smith, L., Elmore, P., & Petry, F. (2017). An approach to explainable deep learning using fuzzy inference. In *Proceedings of the SPIE 10207, Next-Generation Analyst V, 102070D*. https://doi.org/10.1117/12.2268001

Brocardo, M. L., Traore, I., Woungang, I., & Obaidat, M. S. (2017). Authorship verification using deep belief network systems. *International Journal of Communication Systems, 30*(12), e3259. https://doi.org/10.1002/dac.3259.

Daneshvar, A., Homayounfar, M., Fadaei Eshkiki, M., & Doshmanziari, E. (2021). Developing a model for performance evaluation of teachers in electronic education systems using adaptive Neuro-Fuzzy Inference System (ANFIS). *Journal of New Approaches in Educational Administration*, *12*(4), 176-190. https://doi.org/10.30495/jedu.2021.22870.4633

Daniels, L, M., & Bulut, O. (2020). Students' perceived usefulness of computerized percentage-only vs. descriptive score Reports: Associations with motivation and grades. *Journal of Computer Assisted Learning*, *36*(2), 199-208. https://doi.org/10.1111/jcal.12398

Divya, N., Preetam, R., Deepthishree, A. M., & Lingamaiah, V. B. (2019). Analysis of road accidents through data mining. In ICCCE 2018: *Proceedings of the International Conference on Communications and Cyber Physical Engineering 2018* (pp. 287-294). Springer Singapore.

Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., & Van Erven, G. (2019) Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. *Journal of Business Research*, *94*, 335–343. https://doi.org/10.1016/j.jbusres.2018.02.012

Göktepe Yıldız, S., & Göktepe Körpeoğlu, S. (2023). Prediction of students’ perceptions of problem-solving skills with a neuro-fuzzy model and hierarchical regression method: A quantitative study. *Education and Information Technologies*. https://doi.org/10.1007/s10639-022-11446-1

García, E., Romero, C., Ventura, S., & Calders, T. (2007). Drawbacks and solutions of applying association rule mining in learning management systems. *In Proceedings of the International Workshop on Applying Data Mining in e-Learning* (ADML 2007), Crete, Greece.

Gu, J., Gao, B., & Zhu, S. (2018). Characterization of bi-domain drosomycin-type antifungal peptides in nematodes: An example of convergent evolution. *Developmental & Comparative Immunology*, *87*, 90-97. https://doi.org/10.1016/j.dci.2018.06.005

Guo, D., Zhang, Y., Xiang, Q., & Li, Z. (2014). Improved radio frequency identification indoor localization method via radial basis function neural network. *Mathematical Problems in Engineering*, 420482. https://doi.org/10.1155/2014/420482

Haghi, S., Aliakbari, M., & Yasini, A. (2023). EFL Teachers' Individual Development Planning Model: A Data-Driven Approach. *Teaching English as a Second Language Quarterly (Formerly Journal of Teaching Language Skills)*, *42*(1), 1-28. doi: 10.22099/tesl.2022.44999.3151

Iatrellis, O., Savvas, I. K., Fitsilis, P. & Gerogiannis, V. C. (2021). A two-phase machine learning approach for predicting student outcomes. *Educational Information Technology*, *26* (1), pp. 69–88, 2021. https://doi.org/10.1007/s10639-020-10260-x

Ibrahim, Z. M., Bader-El-Den, M., & Cocea, M. (2019). *Mining unit feedback to explore students' learning experiences. In Advances in Computational Intelligence Systems: Contributions Presented at the 18th UK Workshop on Computational Intelligence*, September 5-7, 2018, Nottingham, UK (pp. 339-350). Springer International Publishing.

Isabona, J., Imoize A. L., Ojo, S., Karunwi, O., Kim, Y., Lee, C-C., & Li, C-T. (2022). Development of a Multilayer Perceptron Neural Network for Optimal Predictive Modeling in Urban Microcellular Radio Environments. *Applied Sciences*, *12*(11), 5713. https://doi.org/10.3390/app12115713

Jin, J., Liu, Y., Ji, P., & Kwong, C. K. (2019). Review on recent advances in information mining from big consumer opinion data for product design. *Journal of Computing and Information Science in Engineering*, *19*(1), 010801. https://doi.org/10.1115/1.4041087

Jo, H.-S., Park, C., Lee, E., Choi, H.K., & Park, J. (2020). Path loss prediction based on machine learning techniques: Principal component analysis, artificial neural network, and Gaussian process. *Sensors*, *20*(7), 1927. https://doi.org/10.3390/s20071927

Keserci, S., Livingston, E., Wan, L., Pico, A. R., & Chacko, G. (2017). Research synergy and drug development: Bright stars in neighboring constellations. *Heliyon*, *3*(11), e00442. https://doi.org/10.1016/j.heliyon.2017.e00442

Khan, N., Haq, I.U., Khan, S.U., Rho, S., Lee, M.Y., & Baik, S.W. (2021). DB-Net: A novel dilated CNN-based multi-step forecasting model for power consumption in integrated local energy systems. *International Journal of Electronic Power Energy Systems*, *133*, 107023. https://doi.org/10.1016/j.ijepes.2021.107023

Kim, D., Yoon, M., Jo, I. H., & Branch, R. M. (2018). Learning analytics to support self-regulated learning in asynchronous online courses: A case study at a women's university in South Korea. *Computers & Education*, *127*, 233-251. https://doi.org/10.1016/j.compedu.2018.08.023

Kovalchuk, V., Demchuk, O., Demchuk, D., & Voitovich, O. (2019). Data mining for a model of irrigation control using weather web services. In *Advances in Computer Science for Engineering and Education 13* (pp. 133-143). Springer International Publishing. https://doi.org/10.1007/978-3-319-91008-6_14

Liu, O. L., Brew, C., Blackmore, J., Gerard, L., Madhok, J., & Linn, M. C. (2014). Automated scoring of constructed-response science items: prospects and obstacles. *Educational Measurement, Issues and Practice*. *33*(2), 19–28. https://doi.org/10.1111/emip.12028

Mahendra, M., Kishore, C., & Prathima, C. (2019). Data Mining Efficiency and Scalability for Smarter Internet of Things. In: *Computational Intelligence and Big Data Analytics. SpringerBriefs in Applied Sciences and Technology*. Springer, Singapore. https://doi.org/10.1007/978-981-13-0544-3_11

Mehdi, R., & Nachouki, M. (2022). A neuro-fuzzy model for predicting and analyzing student graduation performance in computing programs. *Educational Information Technology, 28*, 2455–2484. https://doi.org/10.1007/s10639-022-11205-2

Miguel-Cruz, A., Aya-Parra, P. A., Rodríguez-Dueñas, W. R., Camelo-Ocampo, A. F., Plata Guao, V. S., Correal O, H. H., & Quiroga-Torres, D. A. (2018, May). Using data mining techniques to determine whether to outsource medical equipment maintenance tasks in real contexts. In *World Congress on Medical Physics and Biomedical Engineering 2018: June 8, 2018, Prague, Czech Republic (Vol. 3)* (pp. 295-298). Singapore: Springer Singapore.

Naresh, C., Bose, P.S.C., & Rao, C.S.P. (2020). Artificial neural networks and adaptive neuro-fuzzy models for predicting WEDM machining responses of Nitinol alloy: a comparative study. *SN Applied Science*, *2*, 314. https://doi.org/10.1007/s42452-020-2083-y

Noreña–P, A., González Muñoz, A., Mosquera-Rendón, J., Botero, K., & Cristancho, M. A. (2018). Colombia an unknown genetic diversity in the era of Big Data. *BMC Genomics, 19*(8), 61-73. https://doi.org/10.1186/s12864-018-5194-8

Ranjan, J., & Malik, K. (2007). Effective educational processes: a data mining approach. *VINE: The journal of information and knowledge management systems*, *37*(4), 502-515. https://doi.org/10.1108/03055720710838551

Roy, S., & Garg, A. (2017). Predicting the academic performance of student using classification techniques. *In 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics* (UPCON) (pp. 568-572). IEEE.

Saa, A.A., Al-Emran, M., Shaalan, K. (2020). Mining Student Information System Records to Predict Students' Academic Performance. In: Hassanien, A., Azar, A., Gaber, T., Bhatnagar, R., F. Tolba, M. (eds), *The International Conference on Advanced Machine Learning Technologies and Applications* (AMLTA2019). AMLTA 2019. Advances in Intelligent Systems and Computing, vol 921. Springer, Cham. https://doi.org/10.1007/978-3-030-14118 9_23

Saini, M. K., & Aggarwal, A. (2018). Detection and diagnosis of induction motor bearing faults using multiwavelet transform and naive Bayes classifier. *International Transactions on Electrical Energy Systems*, *28*(8), e2577. https://doi.org/10.1002/etep.2577

Shermis, M. D. (2015). Contrasting state-of-the-art in the machine scoring of short-form constructed responses. *Educational Assessment, 20*(1), 46–65. https://doi.org/10.1080/10627197.2015.997617

Son, L.H., & Fujita, H. (2019). Neural-fuzzy with representative sets for prediction of student performance. *Applied Intelligence, 49*, 172–187 (2019). https://doi.org/10.1007/s10489-018-1262-7

Sultana, J., Rani, M.U., & Farquad, M. (2019). Student's performance prediction using deep learning and data mining methods. *International Journal of Recent Technology Engineering, 8*(1S4), 1018–1021.

Talpur, N., Abdulkadir, S. J., Alhussian, H., Hasan, M. H., Aziz, N., & Bamhdi, A. (2023). Deep Neuro-Fuzzy System application trends, challenges, and future perspectives: A systematic survey. *Artificial intelligence review, 56*(2), 865-913. https://doi.org/10.1007/s10462-022-10188-3

Tataria, H., Haneda, K., Molisch, A.F., Shafi, M., & Tufvesson, F. (2021). Standardization of Propagation Models for Terrestrial Cellular Systems: A Historical Perspective. *International Journal of Wireless Information Networks, 28*, 20–44. https://doi.org/10.1007/s10776-020 00500-9

Tofighi, S., & Ahmadi Safa, M. (2023). Fairness in Classroom Language Assessment from EFL Teachers' Perspective. *Teaching English as a Second Language Quarterly (Formerly Journal of Teaching Language Skills), 42*(2), 81-110. doi: 10.22099/tesl.2023.46825.3173

Tomasevic, N., Gvozdenovic, N., & Vranes, S. (2020). An overview and comparison of supervised data mining techniques for student exam performance prediction. *Computer Education*, *143*, 103676. https://doi.org/10.1016/j.compedu.2019.103676

Uskov, V. L., Bakken, J. P., Byerly, A., & Shah, A. (2019). *Machine learning-based predictive analytics of student academic performance in STEM education*, in Proc. IEEE Global Engineering Education Conf. (EDUCON), Dubai, United Arab Emirates, 2019, pp. 1370 1376.

Van der Vleuten, C. P. M., Norman, G. R., & De Graaff, E. (1991). Pitfalls in the pursuit of objectivity: issues of reliability. *Medical education*, *25*(2), 110-118.

Virk, A., Joshi, A., Mahajan, R., & Singh, T. (2020). The power of subjectivity in competency-based assessment. *Journal of postgraduate medicine*, *66*(4), 200–205. https://doi.org/10.4103/jpgm.JPGM_591_20

Yang, F., & Li, F. W. (2018). Study on student performance estimation, student progress analysis, and student potential prediction based on data mining. *Computers & Education*, *123*, 97-108. https://doi.org/10.1016/j.compedu.2018.04.006

Yin, P., Dou, G., Lin, X., & Liu, L. (2020). A hybrid method for forecasting new product sales based on fuzzy clustering and deep learning. *Kybernetes*, *49*(12):3099–3118. https://doi.org/10.1108/K-10-2019-0688Zuo, Y., Kajikawa, Y., & Mori, J. (2016). Extraction of business relationships in supply networks using statistical learning theory. *Heliyon*, *2*(6), e00123. https://doi.org/10.1016/j.heliyon.2016.e00123

Appendix
The data corpus of the study

	Mid	Quiz	Final	Participation	Bonus
S1	22	23	25	3	0
S2	27	38	34	0	0
S3	25	19	35	1/75	0
S4	33	32	34	3/5	1/5
S5	33	47	35	3/5	0
S6	33	51	38	3/5	0
S7	32	49	35	3/25	0
S8	29	34	24	3/5	0
S9	31	36	39	3/25	0
S10	34	47	36	3/5	1
S11	36	55	40	3/5	2
S12	34	40	37	3/5	1/75
S13	28	34	31	3	0/75
S14	34	46	33	3/5	0
S15	38	51	40	3/5	0
S16	32	45	42	3/25	0
S17	29	17	26	0	0
S18	26	43	34	3/25	0
S19	0	30	45	0/5	0
S20	29	35	36	3/75	0
S21	33	30	46	4	0
S22	34	42	49	3/75	0
S23	33	38	40	3	0
S24	31	34	41	3	0
S25	32	39	42	3/75	0
S26	26	38	40	4	0
S27	29	34	46	4	0
S28	31	22	3	1/5	0
S29	32	50	49	4	0
S30	30	41	42	4	0
S31	30	49	44	3/75	0
S32	30	50	45	3/5	0
S33	32	50	46	4	0
S34	30	43	40	3/75	0
S35	30	44	43	4	0
S36	32	46	41	3/5	0
S37	32	46	35	3/5	0
S38	36	47	45	4	0
S39	32	49	39	2/5	0
S40	27	47	40	3/5	0
S41	33	48	43	3/25	0
S42	25	35	40	3/75	0
S43	29	36	41	3	0
S44	33	47	38	4	0
S45	31	48	42	4	0
S46	27	48	39	3/5	0
S47	22	49	39	3	0
S48	30	48	42	4	0
S49	24	48	36	3/75	0
S50	29	49	39	3/5	0
S51	26	28	40	1/75	0
S52	33	30	47	4	0
S53	28	32	49	1/75	0
S54	35	42	49	5	0
S55	37	43	52	5	0
S56	32	32	48	4/5	0
S57	34	41	48	5	0
S58	19	26	41	3	0
S59	38	49	55	5	0
S60	30	44	40	3/25	0
S61	19	36	35	5	0
S62	21	28	41	0/75	0
S63	23	31	44	2	0
S64	15	29	36	3/5	0
S65	24	29	31	4	0
S66	28	36	45	2/5	0