

## The Use of Expert Systems in Individualized Online Exams

**Irfan Simsek**

*Istanbul University - Cerrahpasa  
irfan@istanbul.edu.tr*

**M. Erdal Balaban**

*Gelisim University  
mebalaban@gelisim.edu.tr*

**Hatice Ergin**

*Istanbul University (Emeritus)  
htcerginn@gmail.com*

### ABSTRACT

With the Expert Examination System developed within the scope of this study, questions can be prepared in accordance with the measurement and evaluation criteria in education, which helps measure the students' actual knowledge levels effectively. It is possible to create exam forms by using these questions. The developed Special Examination System records the data of the student and the student's interaction with the system and determines the questions to be used to measure the knowledge level of the student and the order of the system in accordance with the information obtained from the student. In addition, what has also been taken into consideration in the development of this system is that it can be developed according to the needs and student characteristics, and that it can be adapted to the current situation. The aim of this study is to develop an individual exam module where the questions are categorized according to cognitive levels, item difficulty and item discrimination are calculated automatically and difficulty levels are adjusted according to the situation of answering the questions of the person during the examination in both classroom and online environments.

**KEYWORDS:** *online exam, expert system, expert exam system, expert agent, computer adaptive test*

### INTRODUCTION

In this age when computer technology penetrates every aspect of human life, it is necessary to take advantage of this technology in online exams. Expert Systems are a branch of the subject of artificial intelligence and are mainly programs that use human knowledge intensively in problem solving at a specialist human level. These systems are computer programs that can do the work of a specialist person in any complex system. These systems can also be classified as computer programs that can-do jobs that demand expertise such as advising, analyzing and classification, and are capable of performing diagnosis and so on. These programs combine expert knowledge and the ability to use the knowledge necessary to solve problems (Kidd, 2012). Alberico & Micco (1990) define expert systems as computer programs that can be used instead of experts in problem solving and decision making.

Expert systems can be expressed as a transition from data processing to information processing. In data processing, the database is processed effectively depending on an algorithm; in information processing on the other hand, the knowledge base consists of rules and facts which are removed without being bound by any algorithm (Haque, 2013). The expert system is a computer program that emulates the behavior of the expert who solves real-life problems in a particular area. The creation of an expert system consists of two steps (Baur, 1991). The first step is to collect the information and methods to be used in problem solving from the expert person or people. The second and final step is to rearrange the information and methods in a structure that was created to be used later. These processes fall into the field of information engineering. The most significant difference between expert systems and classical computer programs is that expert systems evaluate the data and the classical computer programs use the data (Kidd, 2012).

It is a fact that the examination modules used in the current e-learning systems in our country do not provide a complete assessment-evaluation facility. This is because, the questions asked are not formed according to the level of questions and it is not checked whether the questions are prepared according to the measurement and evaluation criteria in education. Therefore, it can be seen that student knowledge cannot be measured in a distinctive way and the results obtained from these exams cannot provide an objective evaluation. Since there is no individualized online examination system to meet these needs, a web based expert examination system has been developed.

The problem question of this research is “Is the expert examination system usable for individualized online exams?” The expert examination system developed for this research was prepared based on the criteria of preparing the questions according to different cognitive levels, determining the difficulty levels according to the item analysis, setting the question types according to the student potential (individually), and the criteria based on testing students in the classroom and individually. For the purpose of the study, answers to the question “What are the opinions of the experts about the use of expert systems in individualized online exams? was sought.

## LITERATURE REVIEW

### Expert Systems

Expert systems are often the ones that perform the tasks performed by the expert of the field. According to Kidd (2012) expert system is a computer program that can solve problems or make suggestions with its expert knowledge and reasoning ability. These programs combine expert knowledge and the ability to use the knowledge necessary to solve problems (Jackson, 1998). Alberico & Micco (1990) define expert systems as computer programs that can be used in place of experts in problem analysis and decision making.

The expert system can normally perform functions that require human expertise or play a supporting role in decision-making. If the person in the decision-making process is an expert, it can consolidate the decisions. People using such programs can reach the expert level in practice thanks to the technical information given by the program even if they are not experts in the subject (Baur, 1991; Pigford & Bauer, 1994).

Expert systems can be expressed as a transition from data processing to information processing. In data processing, the database is effectively processed based on an algorithm, while in the information processing, the information base in the processing is effectively processed by the rules and facts extracted from an algorithm without being bound to it (Durkin & Durkin, 1994; Jackson, 1998).

The expert system is a computer program that emulates the behavior of the expert who solves real-life problems in a particular area. The creation of an expert system consists of two steps (Baur, 1991). The first step is to collect the information and methods to be used in problem solving from the expert person or people. The second and final step is to rearrange the information and methods in a structure that was created to be used later. These processes fall into the field of knowledge engineering. The most important difference that distinguishes the expert systems from the classical computer programs is that the expert systems evaluate the data, while the classical computer programs use the data.

An expert system is ready to be available 24 hours a day, unlike a specialist who has needs such as sleeping, resting, and having a holiday. It is also impossible to have a specialist in many cases, but it is possible to have many expert systems. Unlike humans, an expert system does not die and take away all its information, the knowledge of an expert system can easily be copied and stored. A computer expert system can always be accepted as at the peak of its performance compared to the expert, because the reliability of the recommendations the experts brings when they are tired or sick can be questioned. In contrast, a well-developed expert system will always generate the best possible recommendations within its limits (Collins, 2018; Ericsson, Hoffman, Kozbelt, & Williams, 2018).

An expert system consists of three components; knowledge base, result extraction machine and user interface. The knowledge base contains the necessary definitions, rules and facts that the system will use for a particular problem. The result extraction machine determines when and how the facts and rules are used to decide, or in other words, to solve the problem. The user interface provides the communication between the user who wants to solve the problem and the problem-solving system, the expert systems (Ericsson et al., 2018).

Expert systems are used instead of experts. Comparison of an expert system with an expert is given in Table 1 (Durkin & Durkin, 1994). It is visible that the expert systems have more outcomes.

**Table 1:** The Comparison of the Expert and Expert System

Factor	Expert	Expert System
Time availability	Work day	Always
Location	Local	Anywhere
Safety	Can't be replaced	Replaceable
The possibility of loss	Yes	No
Performance	Variable	Consistent
Speed	Variable	Consistent
Cost	High	(usually faster) Affordable

The most important feature that distinguishes expert systems from other computer programs is that the expert systems are in a modular structure and to insert the knowledge for the solution of the problem in a portable knowledge base rather than source codes.

Nabiyev, Çakiroğlu, Karal, Erümit, & Çebi (2016) explains the internal structure of expert systems briefly as follows;

- *The knowledge base* is the unit that allows the creation of new information from the information that is being stored in a special way. Knowledge base is the cornerstone of an expert system. It can also be defined as the brain of the expert.
- *The task of the expert* is to provide new information to the system. This process of gaining knowledge is often so complex to be done by an ordinary person.
- *The information gaining module* creates an interface between the expert and the knowledge base. It is responsible for taking the information given by the expert in a format appropriate to the knowledge base and placing it in the knowledge base.
- *The database* is in the form of a traditional relational database. Basic objects or properties are stored in this database.
- *The logical conclusion mechanism* is the unit responsible for conducting research on the knowledge base and making logical conclusions using the database. By referring to the rules and facts, they understand what they mean and perform the reasoning function.
- *The auxiliary interpretation module* helps the logical conclusion mechanism to produce logical results.
- *The user interface* acts as an adapter for the user to communicate with the system. It puts the user questions into a format that the mechanism of conclusion can understand.
- *The user* is a non-expert person. This person does not have to know about the internal structure of the system.
- *Description Unit*; explains how the judgments are made. In addition, if the user asks some questions in the communication process and if the user wants to know why this question is asked, the explanation unit makes the necessary explanation.

One of the general techniques of expert systems is character and word matching technique. Any changes to the knowledge base of a developed system will not affect the entire system. There are decision units to decide on their own. These and similar features make expert systems different from the other programs.

#### Areas Where the Expert Systems are Used

- **Interpretation:** Identification of states from sensor transmitters.  
*Area of Use:* Voice recognition, Image Analysis, Review
- **Forecast:** Extraction of similar results to given cases  
*Area of Use:* Weather forecast, Grain Estimation
- **Diagnosis:** Detection of system disorders according to observation results  
*Area of Use:* Medicine, Electronics
- **Design:** Object design under limited conditions  
*Area of Use:* Circuit Drawing
- **Planning:** Designing operations  
*Area of Use:* Automatic programming, Military planning
- **Imaging:** Comparison of observations to plan sensitivities  
*Area of Use:* Regulation of Nuclear Power Plants and Cost Management
- **Debugging:** Delivering deficiencies that cause errors  
*Area of Use:* Computer Software

- **Repair:** Implementation of a defined management plan  
*Area of Use:* Automobile, Computer
- **Education:** Determining and correcting student behavior  
*Area of Use:* Consultation, Improvement, Treatment
- **Control:** Interpretation, estimation, repair and monitoring of System Behavior  
*Area of Use:* Air Traffic Control, War Control

### **Individualized Testing in a Computerized Environment**

We see that the individualized first test application was developed by Alfred Binet during the development of the Alfred-Binet IQ test. This work of Binet carries all of the features that an individual test has to carry: a pool of pre-prepared items, questions categorized by difficulty levels, the starting choice, a predefined scoring system, a rule for selecting questions from the pool and a predefined termination rule. Although this first application is very simple in comparison to current applications, it can be said that this first application is the basis for today's tests.

Later in the 1950s, there were some studies, but the progress was not possible. In the 1960s, Fred Lord, from the Educational Testing Service, had important work on this subject. The main idea of Lord, (2012) was “a test with a fixed number of questions is not particularly suitable for students with low and high ability levels. If appropriate questions are used, the same test can be performed in less time without any loss of information.” After that, it was developed by American Naval Forces (Eggen & Straetmans, 2000; Kingsbury & Weiss, 1983; Olsen, Maynes, Slawson, & Ho, 1989).

For the individualized tests to be effective, it is required that they do simultaneous calculations and speedy selection of the questions from the database, and the speedy selection of the next question, thus, the full implementation of such an application was accompanied by the developments in computer technology and decline in the prices in the 1970s. Thus, the concept of “Computer Adaptive Individualized Testing (CAT)” has emerged (Çıkırıkçı-Demirtaşlı, 1999).

### **Properties of Tests Individualized in a Computerized Environments**

Nowadays, individualized tests in a computerized environment are mainly used in TOEFL and GRE tests, but also in many other areas. Those who take these tests do their tests on a computer and answer questions at the computer. The computer does more than just offering a multi-media tool with advanced graphics, sound etc. here. Instead of applying a predetermined set of questions in sequence, the computer selects the questions according to the course of the test. Because asking questions beyond the level of the students (very difficult) or asking questions that are far below the level of ability (very easy) of the individual does not provide much information. In addition, asking questions that are not appropriate to the individual's ability level may cause boredom (Rana, Greenwood, Fox-Turnbull, & Wise, 2018).

To maximize the sensitivity of the exam for gathering information about what the individual knows, the questions to be asked are carefully selected based on the previous questions. According to the point of view of the testee, the difficulty of the examination is shaped according to their abilities. For example; a person who is going to take the test can be tested with a more difficult question if he / she has performed well in a medium difficulty item; or if they performed poorly, they could be tested with a simpler question. Compared to the Static Multiple Choice Tests experienced by almost everyone, computer-assisted tests with a series of invariant items applied to each exam require less items to achieve equal and complete scores (Kingsbury & Weiss, 1983; Thissen, Mislevy, & Wainer, 2000).

Tests applied to the individual in the computerized environment apply the following steps: first pulling out questions from the pool to the level of difficulty appropriate to the level of the individual. Then, throughout the test, continuous adjustment is made at the ability level during the application in order to make the individual's ability estimation in the best possible way. These adjustments are dependent on the performance of the individual until he or she reaches the question. If a question is answered incorrectly, the next question will be easier (point value lower); if the answer is correct, the next question is more difficult (higher score value). The important point is that since the applied questions are applied to hundreds or even thousands of people before being added to the pool, the parameters (difficulty, discrimination index, and frequency factor) are known and the questions to be applied are selected based on these parameters (Thissen et al., 2000).

The progress of such a test application can be summarized in the following steps (Çıkırıkçı-Demirtaşlı, 1999; Hambleton & Swaminathan, 2013; Parshall, Harnes, Davey, & Pashley, 2009):

1. An estimate is made about the individual's ability level. It is a good approach to choose a medium-level question here, because the first one is the average difficulty, as we do not know anything about the level of ability of the individual. The item pool is grouped by subjects and / or levels of difficulty and applied to hundreds or even thousands of people and includes well-defined questions. A good pool of material is expected to contain a large number of questions, so the likelihood of applying the same questions is reduced. Over time, it is possible to add new questions to the pool and remove questions from the pool.
2. Among the questions in the item pool, the most appropriate one is selected. If the correct answer to this question is given, the more difficult one, and if the wrong answer is given, then the easier question will be asked. Only correct or only incorrect answers cannot be used in estimating skills.
3. A new ability estimation is made about the individual. Although there are different mathematical methods for estimating ability, two methods are used: Maximum Likelihood and Bayesian. While the Maximum Likelihood method is based on selecting the question that gives the most information about the individual, Bayesian chooses the question that will minimize the next ability estimation range of the individual.
4. Steps 2 and 3 are repeated until a preset stop criterion is met. The criteria used to finish the test can be going below a standard number of errors, a predetermined number of questions or time limit. The difficulty level of the questions asked to the individual is effective in quantifying the questions into scores. While the score value of difficult questions is low, the score value of easy questions is lower. The scores of the individuals who answered the same number of questions vary according to the number of questions they answered.

The important advantages of the individualized test application can be listed as follows (Durkin & Durkin, 1994; Ericsson et al., 2018):

- Shorten the application time of the test,
- Each student receives a test of his or her own level,
- Increases security,
- Test results can be evaluated instantly,
- The test can be given at any time,
- The test eliminates the use of paper,
- It is easy to remove unwanted questions from the question pool,
- Expands test standardization,
- Reduces test proctoring time,
- Increases flexibility in question selection.

In addition to these advantages, there are some disadvantages as well (Çıkrıkçı-Demirtaşlı, 1999; Hambleton & Swaminathan, 2013; van der Linden & van Krimpen-Stoop, 2003):

- Not applicable for every subject or skill,
- Limits and cost of computer hardware may be a problem,
- Fear of computer may cause problems in such applications,
- The fact that the measured ability is not a one-dimensional requirement is not met,
- The need to have a large pool of question items

### Item Response Theory

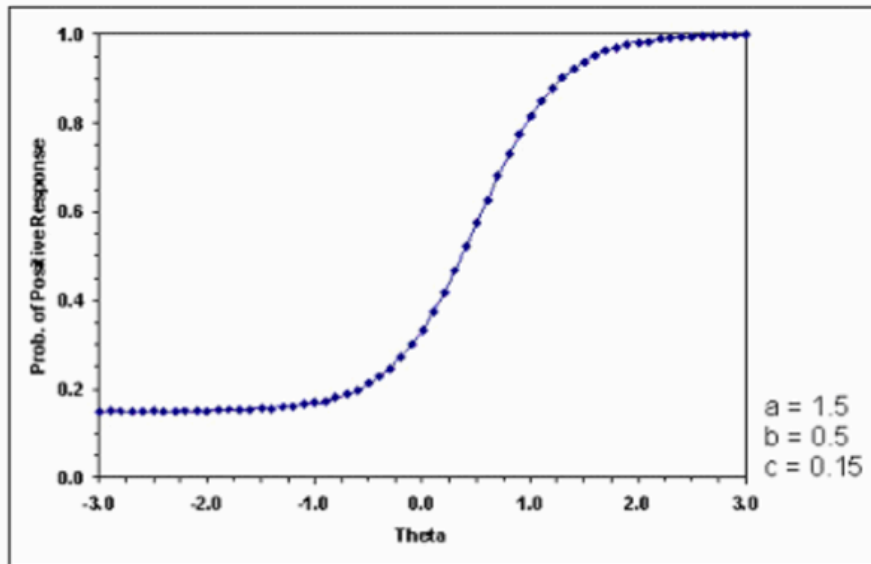
The most appropriate mathematical theory that can be used for individualized tests in the computerized environment is the Item Response Theory. The most important aspect of this theory is that it makes ability estimations independent from the applied questions. Even if individuals are tested with different questions, Item Response Theory provides a standard framework for predicting the abilities of different individuals (Hambleton & Swaminathan, 2013; Parshall et al., 2009; Yenil, 1995). Even if two different tests with different questions are applied to the same individual, the predicted ability level is not different. The theory has the assumption that there is only one factor explaining every ability, but it is clear that such an assumption cannot be easily achieved. For example; if we consider the English language proficiency, we cannot say that it consists of a single factor, because it includes factors such as reading comprehension, vocabulary, listening and writing (Embretson & Reise, 2013; Hambleton & Swaminathan, 2013; Lord, 2012).

Item Response Theory suggests three different models consisting of one parameter, two parameters and three-parameter models. One parameter model correlates with the item difficulty parameter ( $b_i$ ) and the level of the individual's ability (Bolt, Cohen, & Wollack, 2001; Crocker & Mazer, 2019). In the two-parameter model, the distinction parameter is added to the item difficulty parameter (distinction power of a question between the



low and high ability groups -  $a_i$ ). In the third model, the chance factor parameter (the probability of low-level individuals correct response -  $c_i$ ) is added to the two-parameter model (Hambleton & Swaminathan, 2013).

The three-parameter logistic model of the Item Response Theory is the most suitable for multiple-choice question types in individualized computerized test applications. In this model, there are three parameters for each question:  $a_i$ ,  $b_i$ , and  $c_i$ .  $a_i$  question is the index of discrimination (distinction power of a question between the low and high ability groups at a certain ability level),  $-b_i$  question represents the degree of difficulty (the level of ability the item provides the most information) and  $-c_i$  represents the forecast index (Lin & Spray, 2000). These parameters are different for each question. An example item characteristics curve graph calculated based on these parameters according to the probability of giving correct answer to the question in one level is shown in Figure 1.



**Figure 1:** Item Characteristics Curve

In this graph, the x-axis (theta) represents the ability levels and y-axis (prob) represents the possibility of correct answers of the problem.  $a_i$  is the slope of the point where the twist point of the graph (prob = 0.5) intersects the x axis. The larger the value of  $a_i$ , the steeper the chart and the more distinctive it will be. The value of  $-b_i$  will move the graph to the left or to the right (the probability of the correct answer to the question will vary for the ability levels, the more to the left the question is the easier the question is; the more right the question goes the more difficult it is).  $c_i$  is the point at which the graph intersects the y-axis, and the ability of those at very low skill levels to make this question, i.e., when a random answer is marked, gives the possibility that the answer is correct. The higher the value, the higher the likelihood of correct answer by the lower ability level students (Çelik, Baran, & Sert, n.d.). The selection and grading of the questions in the test application are done according to these parameters. Among the weaknesses of the Item Response Theory are the need for large sample groups, the complexity of the theory and the difficulty of interpreting the results, the need for special software, and a number of assumptions that their provision is more difficult than the classical test theory (Bolt et al., 2001; Hambleton & Swaminathan, 2013).

There are other methods besides Item Response Theory for individualized tests. Lawrence Rudner developed the Measurement Decision Theory to be used to classify individuals. Sequential Probability Ratio Test is an approach used to classify individuals according to their competencies (Kingsbury & Weiss, 1983; Piton-Gonçalves & Aluisio, 2012).

### Use of Individualized Tests in Education

Individualized tests in computerized environment are currently used in GMAT (Graduate Management Admission Test), GRE (Graduate Record Examination), TOEFL (Test of English as a Foreign Language) applications. In these tests, individuals are asked questions from a pool of previously prepared questions with known parameters. Each question is determined according to the individual's performance in previous questions. These tests are carried out by prior appointment in certain test centers.

For the application of quantitative and verbal parts in GRE, 58 questions are used and the application lasts 75 minutes. The TOEFL application takes 165 to 210 minutes, but this also includes the application of the listening

and writing parts. Similarly, the GMAT application lasts 210 minutes. These exams provide high reliability information about the level of ability they measure in shorter time using fewer questions (GRE & TOEFL Official websites, 2018).

In a study which included the comparison of individualized tests on computerized environments to pen and paper exams according to validity and reliability (Lord, 2012), the following conclusions were found: individualized tests gave high reliable results than the pen and paper tests with less than half the number of questions used; in large-scale samples, the validity of individualized tests was found to be better. In summary, this study provides verbal aptitude predictions with high reliability when compared to the same length pen and paper tests in 15 item individualized tests.

There is an extensive literature on individualized testing in computerized environments practices abroad. World leading individualized test applications in computerized environment include:(Olsen et al., 1989; Parshall et al., 2009; Thissen et al., 2000):

- Adaptive Matrices Test (AMT),
- ASCP (American Society of Clinical Pathologists-Board of Registry Certification Examinations),
- ASVAB (the Armed Services Vocational Aptitude Test Battery),
- CAT of Written English for Spanish Speakers,
- BULATS (Business Language Testing Service) Computer Test,
- CATE (Computerized Adaptive Test of English),
- COMPASS series of tests from ACT, GMAT (Graduate Management Admission Test,
- GRE (Graduate Record Examination),
- LPCAT (Learning Potential CAT),
- MAP (Measures of Academic Progress),
- Microsoft Certified Professional exams,
- NAPLEX (North American Pharmacist Licensure Examination),
- NCLEX (National Council Licensure Examinations)
- STAR Math, Reading, and Early Literacy tests

Considering the studies on individualized test applications in our country, İşeri (2002) came to the following conclusions by measuring the mathematics achievement with the computerized individualized test by using the question bank compiled from the mathematics parts of the Secondary Education Institutions Selection Placement Exam and of the Private Schools Exam: individualized tests in the computerized environment reliably measure the ability level using less questions; if students were allowed to control their answers, students at high level of talent were estimated as low talent levels; The Bayesian method provided better estimation, and both fixed number and stationary false stop methods yielded good results.

Weller (2019) compared the individualized tests on computerized environments with the pen and paper tests in his study. In this study, he has developed and implemented a test consisting of past University Entrance Exam questions. 14 questions were applied in computerized environment against 50 questions in a pen and paper test. At the end of the study, the test showed significant decreases in application time and number of questions. There was no significant difference between ability estimates.

Şahin & Kışla, (2013) in their study examining the compatibility of the Private Schools Exam data with the Item Response Theory, found that one-dimensionality, local independence, low chance of success were met and that the Item Response Theory was appropriate for evaluating the exam.

## METHODOLOGY

The research was designed in a qualitative design. Semi-structured interview form was used as data collection tool. Interviewing is one of the most widely used data collection tools in qualitative research. Creswell (2002) defined the interview as “a process of mutual and interactive communication based on the way of asking and answering questions for a predetermined and serious purpose” (Yildirim & Şimşek, 2005).

The working group comprised of the faculty members of the Istanbul University Hasan Âli Yücel Faculty of Education, Teaching Principles and Methods, and eight assessment and evaluation experts (3 males, 5 females). At the same time, these faculty members (1 Professor, 3 Associate Professor, 4 Assistant Professors) are experts in the field of assessment and evaluation, who have been teaching Assessment and Evaluation and Teaching Principles and Methods for at least five years. In order to obtain the expert opinion in the preparation of the exam form and in the interpretation of the results of the evaluation, objective sampling has been accepted. In this type

of sampling, it is aimed that the researcher believes that he / she will find answers to their problems and that, in accordance with the objectives of the research, only one part of the whole body is observed.

The main features of this system are as follows;

- It records the data of the student as a result of the interaction of the student with the system,
- In the light of the obtained data, it determines the questions to be used in measuring the level of knowledge of the student and the order of operation of the system,
- The system can improve itself according to the changing needs and student structures and adapt itself to the current situation.
- Questions are categorized according to cognitive fields and goals and outcomes,
- Since the content validity measurement is performed by experts who experts in evaluation are, each student has the same question level, but the program can ask different questions.
- Content validity, item difficulty and item differentiation has been measured for each cognitive level for exams which will not be performed online, thus serves as a quality question bank,
- An infrastructure to perform a personalized examination is developed and a real assessment can be done.
- It helps prepare the exam according in accordance with the test plan steps,
- By the help of text editor, it allows for alternative question techniques to be used in the questions and choices.
- While preparing the question and the exam form, the expert agent helps to prepare them according to the educational measurement and evaluation criteria.

## FINDINGS

All the participants have confirmed the necessity of such a system when asked ***“Is there a need for such an expert examination system when preparing the exam form? If so, why?”*** two of the participants answered as follows:

*“I think it is absolutely necessary to have such a system when preparing the exam form. Although teachers have prepared samples of questions having taken assessment and evaluation courses, we see that sometimes people do not internalize assessment and evaluation criteria since they do not have control system (feedback) and sometimes they prepare questionnaires filled with errors without being aware of them. This system gives feedback to the teacher at every step and ensures that the questions are prepared in accordance with the assessment and evaluation criteria. In addition, getting feedback from 3 different experts related to each question further strengthens the system. Thus, the teacher's awareness increases his knowledge and skills.”*

*“There is a need. It will make assessment and evaluation process more reliable, valid and objective. In addition, it will be possible to ask questions from all levels and it will be possible to differentiate between the ones that learn and the ones that do not learn. These transactions will not be left to chance. It will eliminate the differences in terms of the quality of preparing questions among teachers who give the same course in an institution. Unfortunately, sometimes a teacher with 1 or 2 questions can be found. It can be seen from some teachers that 90% of the class has 90 or more points. There is a need for this system in order to prevent the preparation of questions in a way that does not comply with the assessment and evaluation criteria like this and to prepare quality questions.”*

To the question ***“Related to this Expert Examination System, in which Expert Agents and individuals interact; What do you think about the cooperation of expert agents and individuals?”*** two-thirds of the participants stated that the teacher's knowledge and skills will improve with a cooperative study.

*“With this cooperation, it will be possible to prepare more quality questions in a shorter time. Moreover, it will be a system that the teacher can trust.”*

*“I think the points that can be overlooked by the expert agent can be easily captured. This will help the person to feel safe when using the system.”*

*“It is easier for the expert to control each problem in terms of assessment and evaluation criteria. The expert examines the questions in more qualitative terms and gives feedback.”*

*“It is not always possible for a teacher to talk to an expert. However, I think it is a chance for this expert to be brought to the teacher's feet as a computer program.”*

All participants answered “yes” to the question ***“Expert Agent interactions; Is it instructive to help the questioners in terms of assessment and evaluation criteria?”***

*“Yes it is instructive. In fact, it is accepted that the people to prepare the questions learned these criteria and graduated, and they know the criteria after graduation. In practice, however, it is often observed that no questions have been prepared in accordance with these criteria. With this system, it*



*will be possible for the them to remember what they have learned, to re-learn what they have forgotten, and to reinforce what they have learned.”*

*“It is absolutely instructive. Because almost every stage has defined criteria, it gives regular feedback. Through this feedback, we prepare questions that are appropriate to the criteria and on the other hand, the criteria are learned.”*

*“As it is an interactive process, it is possible to say that it will contribute to this direction.”*

*“It will enable the teacher to develop as it gives feedback to the teacher in all stages of preparing questions and creating test forms. “*

Some answers of the participants to the question **“Expert agent in the Expert Exam System, what do you think about the facilitation of experts’ task by checking if each question is in compliance with the criteria for assessment and evaluating?”** are as:

*“It is easier for the expert to control each problem in terms of assessment and evaluation criteria. The expert examines the questions in more qualitative terms and gives feedback.”*

*“The fact that the Expert Examination System checks the compliance of each question with the criteria for preparing an assessment and evaluation question ensures that the expert does not spend time and energy on the quality of the questions. “*

*“If the Expert Agent is programmed very well, it will make the teacher's job much easier.”*

Some of the answers given by the participants to the question **“What is your criticism towards such an Expert Examination System when you were preparing an exam form, what are the challenges?”** are given below:

*“I believe that it is very useful in the long run, even though it is a bit time-consuming until the question bank is created at the beginning. There may be disruptions if the experts do not provide feedback during the specified time. There may be difficulties in finding an expert.”*

*“I believe that when the system is well-structured, it will provide a lot of convenience and benefit on the contrary.”*

*“Anyway, if teachers believe that there are questions in the question bank and whenever they want to mix them they can do so, it can push teachers to laziness about creating new questions. I have concerns about getting past with other teachers' questions.”*

*“Criticism may occur when the system is applied rather than to exam questions and forms. In the current framework, it is a system consistent with its objectives. However, new and unexpected situations encountered in the application will allow for further development of the system.”*

Some of the answers given by the participants to the question **“What are your suggestions for such an Expert Exam System when preparing the exam form?”** are given below:

*“Teachers preparing questions may be resistant to implementing this system. Methods and strategies to reduce this resistance and encourage teachers to use this system can be developed. Because this system is a system that requires constant renewal. However, teachers can ask for the same questions for 10 years. For this reason, it is recommended that the teachers use the system together in the department.”*

*“Before the application of this system, the teachers who will prepare questions in this system can be given a training on the assessment and evaluation criteria. Thus, negative feedback and loss of time can be prevented. I can especially recommend that branch teachers in schools use this system in common.”*

*“I can suggest that the Expert Exam System is a structure where teachers can create their own question banks within the system and share the questions they want.”*

## CONCLUSIONS

In the interview with the Assessment and Evaluation Expert Instructors, they stated that such an Expert Examination System is required in preparing the exam form, and that even though the teachers have the knowledge of assessment and evaluation, when they do not have feedback, they sometimes prepare the wrong exam forms without being aware. The feedback from three different field experts strengthens the system. Quality questions can be prepared in a short time by means of a collaborative study by expert agents and individuals. It will also make the person feel safe when using the system. At the same time, it will help remember the forgotten information and reinforce the existing information. As there is a regular feedback, it is instructive in terms of assessment and evaluation criteria for the question preparing individuals.

The Expert Agent in the individualized online Expert Exam System will enable the expert to facilitate the task of the individual by checking whether each question meets the criteria for preparing an assessment and evaluation, allowing the expert to provide feedback to the questions in a more qualitative way. The criticism about the Expert Examination System is that it does not appear to be useful until just a sufficient number of questions are piled at the beginning of the question bank, and that it will create malfunctioning when the system does not give feedback

during the specified time. There are also criticisms that the ready-made question bank could lead the teachers to laziness about preparing new questions. Although the questioners might resist to use such a system at the beginning, the resistance will be reduced, especially after a sufficient number of questions are piled in the question bank. Besides, it is clear that negative feedback will decrease if the teachers are given a training about the criteria for preparing assessment and evaluation. This system is especially important for the use of branch teachers and to create a common question bank.

Thanks to the Expert Examination System which the Expert Agents and individuals interact with, a quality question bank can be created in accordance with the assessment and evaluation criteria in education. The system has an infrastructure that can perform individual tests. In addition, after the online exams are conducted, an individualized examination can be made as the questions' item analysis are performed. It is also important that the developed Expert Examination System is a system that not only facilitates the preparation of an effective question in accordance with the assessment and evaluation criteria but also eliminates the deficiency of knowledge of the person preparing the questions, thus improves them.

Some of the rules defined by the Expert Agent are as follows:

- When preparing a multiple choice question; to select the number of options it determines 3 choices for classes 1-3; 4 choices for classes 4-8. and 5 choices for classes 9-12.
- Warns the person who prepared the question in case of using “none” or “all” in the correct answer or in or in the distractors.
- Warns the person who is preparing a question about the choices should be given in order if they contain dates or numbers.
- In the case of the matching question type, it is recommended that the premiss should be at least 6 and maximum 15, and that the answer list should be at least 3 more than the premiss.
- When preparing a question in the fill in the gap question type; the question is not to use more than one gap, the beginning of the sentence should not start with a blank, and the sentence should not end in auxiliary verb.
- It helps the person who prepares questions with sample questions and questions about what to look for while preparing a question according to the chosen level of cognitive domain.
- Warns the person who prepared the test that a test form can be used with up to three different types of questions, at the same time matching question type, true / false and gap-filling question type trio should not be used together in the form of an exam form.
- In the form of an exam, there should be questions in each cognitive level and the distribution should be balanced.
- Calculates and shows the number and distribution of questions according to content in the developed exam form.
- Calculates the item analysis and creates question cards after the online exam.
- Calculates the success level of the students after the examination. It also calculates the success level and percentage of the questions asked in the exam according to the cognitive levels.

“Expert Examination System”; as the questions are grouped and stored by the question categories, the teacher has the opportunity to make a real assessment and evaluation since he can access the questions related to the question type at every level while preparing the exam form. As the question category in the “Expert Exam System” is a dynamic structure; when preparing a question, people who do not want to prepare the questions according to Bloom's taxonomy or who want to prepare according to the revised new taxonomy can easily determine the difficulty levels. “Expert Examination System”; unlike the online exam modules included in other e-learning systems sends the questions uploaded to the system to be checked by at least three field experts before signing up to the system. This check is made to evaluate whether the questions to be stored into the system are appropriate for the category in which they are intended to be prepared. The questions are recorded in the system after at least two of these three experts have approved.

In applications developed for commonly used learning management systems such as Moodle, Blackboard and other online exams; it is observed that the exam questions are not prepared according to the assessment and evaluation criteria and the questions are not categorized according to cognitive levels. The “Expert Examination System” has been developed since the questions are not prepared according to cognitive levels and in order to contribute to the elimination of this deficiency, and because preparation of questions is left at teachers' goodwill. “Expert Examination System” has been developed in an object-oriented structure using open source PHP programming language and MySQL database so that it can run smoothly on every platform. Because the system is prepared on the internet, people who want to use this system can access the system with modern web browsers and use the system without any problem. In addition, web service infrastructure has been prepared in order to draw courses, lecturers and students from the automation systems used in the institution.

After the student has passed the exam the “Expert Examination System” provides the students with information about how many correct answers the student made in the exam, how many questions he answered incorrectly, and how many questions he has not answered at all. In addition to this, it also provides information about how many questions are asked in what question category, how many of them could be answered and how many of them are not answered. According to these results, the teacher can see the student's deficiency and direct the student to complete the deficiency. The “Expert Examination System” provides teachers with the opportunity to prepare questions with a high degree of validity and reliability by reducing the errors in preparation to the least. In this way, high validity and reliability are ensured.

#### ACKNOWLEDGEMENT

This study was produced from a doctoral thesis “Development of a Web Based Expert Examination System According to Measurement and Evaluation Criteria in Education”

#### REFERENCES

- Alberico, R., & Micco, M. (1990). Expert systems: for reference and information retrieval. Retrieved from <http://cds.cern.ch/record/211756>
- Baur, G. (1991). Expert Systems for Business: Concepts and Applications. Retrieved from <https://dl.acm.org/citation.cfm?id=574761>
- Bolt, D. M., Cohen, A. S., & Wollack, J. A. (2001). A Mixture Item Response Model for Multiple-Choice Data. *Journal of Educational and Behavioral Statistics*, 26(4), 381–409. <https://doi.org/10.3102/10769986026004381>
- Çelik, S., Baran, E., & Sert, O. (n.d.). The Affordances of Mobile-App Supported Teacher Observations for Peer Feedback. *International Journal of Mobile and Blended Learning*.
- Çıkrıkçı-Demirtaşlı, N. (1999). Psikometride yeni ufuklar: Bilgisayar ortamında bireye uyarlanmış test. *Türk Psikoloji Bülteni*, 5(13), 31–36.
- Collins, H. M. (2018). Expert Systems, Artificial Intelligence and the Behavioural Co-ordinates of Skill\*. In *The Question of Artificial Intelligence* (pp. 258–281). Routledge. <https://doi.org/10.4324/9780429505331-6>
- Creswell, J. W. (2002). *Educational research: Planning, conducting, and evaluating quantitative*. Prentice Hall Upper Saddle River, NJ.
- Crocker, S. G., & Mazer, J. P. (2019). Associations among community college students' technology apprehension and achievement emotions in developmental education courses. *Technology, Pedagogy and Education*, 1–16.
- Durkin, J., & Durkin, J. (1994). Expert systems: design and development. Retrieved from <https://eetyfiyw3hz08.storage.googleapis.com/EgNhdLLye8nnSiK0DK08.pdf>
- Eggen, T. J. H. M., & Straetmans, G. J. J. M. (2000). Computerized Adaptive Testing for Classifying Examinees into three Categories. *Educational and Psychological Measurement*, 60(5), 713–734. <https://doi.org/10.1177/00131640021970862>
- Embretson, S., & Reise, S. (2013). Item response theory. Retrieved from <https://content.taylorfrancis.com/books/download?dac=C2009-0-10676-2&isbn=9781135681470&format=googlePreviewPdf>
- Ericsson, K., Hoffman, R., Kozbelt, A., & Williams, A. (2018). The Cambridge handbook of expertise and expert performance. Retrieved from <https://www.google.com/books?hl=tr&lr=&id=HM5TDwAAQBAJ&oi=fnd&pg=PR12&dq=%22expert+systems%22&ots=bbYibUuvy8&sig=cBg20ElGcwp8OqpPVJLnbC-0Tz8>
- Hambleton, R., & Swaminathan, H. (2013). Item response theory: Principles and applications. Retrieved from [https://www.google.com/books?hl=tr&lr=&id=dUbwCAAQBAJ&oi=fnd&pg=PA33&dq=The+Fit+of+One-,+Two-+and+Three-Parameter+Models+of+Item+Response+Theory+to+the+Student+Selection+Test+of+the+Student+Selection+and+Placement+Center&ots=jtGKMqr0VF&sig=S-YhrPl\\_mz09pBXSt2DmFDv2nM](https://www.google.com/books?hl=tr&lr=&id=dUbwCAAQBAJ&oi=fnd&pg=PA33&dq=The+Fit+of+One-,+Two-+and+Three-Parameter+Models+of+Item+Response+Theory+to+the+Student+Selection+Test+of+the+Student+Selection+and+Placement+Center&ots=jtGKMqr0VF&sig=S-YhrPl_mz09pBXSt2DmFDv2nM)
- Haque, M. (2013). Application of artificial intelligence in construction management. Retrieved from [https://scholar.google.com.tr/scholar?hl=tr&as\\_sdt=0%2C5&q=%22Application+of+Artificial+Intelligence+In+Construction+Management%22&btnG=](https://scholar.google.com.tr/scholar?hl=tr&as_sdt=0%2C5&q=%22Application+of+Artificial+Intelligence+In+Construction+Management%22&btnG=)
- İşeri, A. İ. (2002). Assessment of Students Mathematics Achievement Through Computer Adaptive Testing Procedures. Middle East Technical University. Department of Secondary Science and ....
- Jackson, P. (1998). Introduction to expert systems. Retrieved from <https://dl.acm.org/citation.cfm?id=521024>
- Kidd, A. (2012). Knowledge acquisition for expert systems: A practical handbook. Retrieved from [https://www.google.com/books?hl=tr&lr=&id=SLsxBQAAQBAJ&oi=fnd&pg=PA1&dq=Knowledge+Acquisition+for+Expert+System&ots=haFONIVUki&sig=1kKQAkzxcF3NXDav\\_oUiSpfqYvk](https://www.google.com/books?hl=tr&lr=&id=SLsxBQAAQBAJ&oi=fnd&pg=PA1&dq=Knowledge+Acquisition+for+Expert+System&ots=haFONIVUki&sig=1kKQAkzxcF3NXDav_oUiSpfqYvk)

- Kingsbury, G. G., & Weiss, D. J. (1983). A comparison of IRT-based adaptive mastery testing and a sequential mastery testing procedure. In *New horizons in testing* (pp. 257–283). Elsevier.
- Lin, C., & Spray, J. (2000). Effects of Item-Selection Criteria on Classification Testing with the Sequential Probability Ratio Test. ACT Research Report Series. Retrieved from <https://eric.ed.gov/?id=ED445066>
- Lord, F. (2012). Applications of item response theory to practical testing problems. Retrieved from <https://content.taylorfrancis.com/books/download?dac=C2007-0-02656-X&isbn=9781136557248&format=googlePreviewPdf>
- Nabiyev, V. V., Çakiroğlu, Ü., Karal, H., Erümit, A. K., & Çebi, A. (2016). Application of Graph Theory in an Intelligent Tutoring System for Solving Mathematical Word Problems. *Eurasia Journal of Mathematics, Science & Technology Education*, 12(4), 687–701.
- Olsen, J. B., Maynes, D. D., Slawson, D., & Ho, K. (1989). Comparisons of Paper-Administered, Computer-Administered and Computerized Adaptive Achievement Tests. *Journal of Educational Computing Research*, 5(3), 311–326. <https://doi.org/10.2190/86RK-76WN-VAJ0-PFA3>
- Parshall, C. G., Harnes, J. C., Davey, T., & Pashley, P. J. (2009). Innovative Items for Computerized Testing. In *Elements of Adaptive Testing* (pp. 215–230). New York, NY: Springer New York. [https://doi.org/10.1007/978-0-387-85461-8\\_11](https://doi.org/10.1007/978-0-387-85461-8_11)
- Pigford, D., & Bauer, G. (1994). Expert Systems for Business; Concepts and Applications. Retrieved from <https://dl.acm.org/citation.cfm?id=561672>
- Piton-Gonçalves, J., & Aluísio, S. M. (2012). An architecture for multidimensional computer adaptive test with educational purposes. In *Proceedings of the 18th Brazilian symposium on Multimedia and the web* (pp. 17–24). ACM.
- Rana, K., Greenwood, J., Fox-Turnbull, W., & Wise, S. (2018). A Shift from Traditional Pedagogy in Nepali Rural Primary Schools? Rural Teachers' Capacity to Reflect ICT Policy in Their Practice. *International Journal of Education and Development Using Information and Communication Technology*.
- Şahin, M., & Kışla, T. (2013). KİŞİSELLEŞTİRİLEBİLİR ÖĞRENME ORTAMLARI: LİTERATÜR İNCELEMESİ.
- Thissen, D., Mislevy, R. J., & Wainer, J. (2000). Computerized adaptive testing: A primer. *Testing Algorithms*, 101–133.
- van der Linden, W. J., & van Krimpen-Stoop, E. M. L. A. (2003). Using response times to detect aberrant responses in computerized adaptive testing. *Psychometrika*, 68(2), 251–265. <https://doi.org/10.1007/BF02294800>
- Weller, J. (2019). Critical Reflection through Personal Pronoun Analysis (Critical Analysis) to Identify and Individualise Teacher Professional Development. *Teacher Development*.
- Yenal, E. (1995). Differential item functioning analysis of the quantitative ability section of the first stage of the university entrance examination in Turkey. *Ortadoğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü. Yayınlanmamış Yüksek Lisans Tezi*.
- Yildirim, A., & Şimşek, H. (2005). Sosyal bilimlerde nitel araştırma yöntemleri. *Ankara: Seçkin Yayıncılık*.