The Measurement of the Student's Basic Knowledge Obtained by Means of Computer Assessments

Gennadiy N. Burlak[1], José Alberto Hernández[1],
Alberto Ochoa[2] and Jaime Muñoz Arteaga[3]

[1] Centro de Investigación en Ingeniería y Ciencias Aplicadas,
Universidad Autónoma del Estado de Morelos,
Av. Universidad 1001 Col. Chamilpa, Cuernavaca, Morelos México, CP. 62209
gburlak@uaem.mx

[2] Universidad Autónoma de Zacatecas
Jardín Juárez 147 Centro Histórico, Zacatecas, México, CP. 98000

[3]Universidad Autónoma de Aguascalientes

Abstract. We discuss the application of the computer technology on assessing the level of basic knowledge for a wide spectrum of subjects. This approach consists in questioning and then evaluating the validity of the students' answers. The employment of the Monte-Carlo technique for estimation of the optimal parameters allows us to develop a new approach to increase the accuracy of the valuation and determination of the stability of the knowledge. As implementation we use the Client-Server technology based on the natural evaluation process, where the students (Clients) are tested by the remote examiner (Server) in online regime. We focus on the Question and Testing Interoperability stage and the possibility of creation of the learning objects. We find that the optimal regime of assessment can be achieved at the specific random generation of tasks (assessment items) in the test. The construction of new adaptive methods and the program realizations of such technology are discussed also.

Keywords: Online assessments, Monte Carlo Method, Client-Server Technology

1 Introduction

The goal of this paper is to consider the evaluation of the level of student's knowledge at computer testing. We discuss the following questions: (i) What is the level of knowledge and how can it be estimated (ii), Why the random numbers must be used, (iii) How the statistically stable conclusions can be obtained, and (iv) Which are the optimum conditions for the evaluation of the student knowledge level.

© A. Gelbukh, S. Suárez. (Eds.) *Received 08/07/06*
Advances in Computer Science and Engineering. *Accepted 03/10/06*
Research in Computing Science 23, 2006, pp. 201-213 *Final version 17/10/06*

This paper is organized as follows. In Sec.1 we discuss the basic strategies and methods of the tests; we present some numerical study of the student's knowledge level. In Sec.2 we present the algorithms and the program implementation of the computer testing. First, we analyze the online testing system (OTS) to assess the students' basic knowledge through online tests implementation with the use of learning objects [12] on the base of Client-Server technology. We focus on the e-learning Question and Test Stage, and the generation of learning objects in XML framework. Our goal is to accomplish the standard IMS Question & Test Interoperability Specification version 1.2.1 [14]. The OTS is not competition for well-known systems similar Blackboard [20] that offers an integrated solution for e-learning. We consider a similar solution as the Questionmark Perception Assessment software [19]. This enables educators to write, manage and report about assessments on a secure manner in Windows and web environments on a secure manner, storing the results for further processing in e.g. Oracle or SQL-Server Databases. The difference with respect to Questionmark is that presented OTS is platform independent, and it can save the database license expenses since storing the results in XML - MySQL databases. In Sec. 3 we present the results of practical implementation, and finally, we discuss and summarize our conclusions.

2 Approach to Methods of the Test

First we discuss the structure of the testing knowledge. Let us assume the knowledge base (KB) as a collection (set) of the coupled questions-answers ($\boldsymbol{x_i}$, $\overline{f}_{ij}$, T) (records). In such a collection $\boldsymbol{x_i}$ is the asked question and $\overline{f}_{ij}$ are proposed answers to this question, $\boldsymbol{i}$ is the current number, j is number of answer in the list of proposed answers, and T is time to test. Here i<=N, j<=N_a , N is number of records in the collection, and N_a is number of proposed answers. To the sake of simplicity we further refer to the experts (teachers) knowledge base as the etalon KB. We assume that answers $\overline{f}_{ij}$ are ordered on the closeness to the correct answer. Then $F_i = \overline{f}_{i1}$ is the correct answer to i-th question. (We note that before the presentation all data undergo to random mixing.) At evaluating, the question is addressed to the student and his answer is compared with the corresponding correct answer. After testing the initial collection is extended with obtained results, e.g., obtained answers and other relevant information (response time, rating, etc.). After test, the initial collection is extended by the obtained answers, and there contains all necessary information for the estimation of the student's knowledge [7].

There are two different types of examination. In first type the student must recognize the correct answer from several predefined variants $\overline{f}_{ij}$ (the closed-ended form of the test). In the second type the system allows student to write his answer freely (the open-ended form of the test). Obviously that in first case (closed-ended test) the problem of verification of the answer can be solved easily. But this problem requires much more efforts in the case of open-ended test. Since the correct answer can be written in various but equivalent forms, the problem of reduction of the answer to the unique referenced form has obvious solutions only for simplest cases. But in general such a problem may turn out quite involved. Here we pay main attention to the closed-ended test case.

We assume that the levels of the knowledge are ordered as follows: the quantity F_i is greater or equal with respect to other answers $F_i >= \overline{f}_{ij}$. We suppose that it is only the unique right answer to every question in the etalon collection [3].

Further we renormalize $\overline{f}_i$ and F_i with some coefficients w_i. With the use $w_i=1/F_i$ the etalon knowledge is represented by a straight line $F_i=1$, while the student knowledge is rewritten as $\overline{f}_{ij}/F_i = f_{ij}$, where $0 \le f_{ij} \le 1$.

In Fig.1 are shown schematically both distributions: the etalon knowledge and the student knowledge. The line at the top represents the level of correct (etalon) knowledge (knowledge of the teacher), and the variant line the student's knowledge. ***X***-axis corresponds to the number of asked question, while in the ***Y***-axis the level of student's answer is postponed.

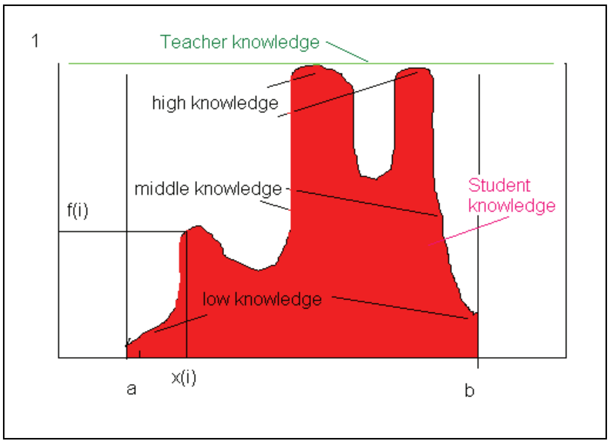


Fig. 1. Distribution of normalized etalon KB (green color) and distribution of student's KB (correct, wrong) in general (red color).

Here N is total number of records (tasks) in the collection, while n is the number of records (questions) in the current test.

The details of the graphic representation f_{ij} depend on the rules used at the examination. In the advanced cases the quantity (proposed answers) f_{ij} may have a value proportional to a closeness of the student answer to the etalon answer F_i. In a practical important case, one may use the following simple rule of the estimation: $f_{ij}=1$ for correct answer, and $f_{ij}=0$ if the answer is incorrect. For a small number of questions n such graphical representation is the histogram. At great number $n>>1$ the discrete picture can be generalized to the case of continuous distribution of the knowledge as

$$dZ = f(x)dx, \tag{1}$$

In this case ***f(x)*** is the local density of knowledge at vicinity of ***x*** and ***x+dx***.

In general the normalized knowledge ***Z*** may be defined as

$$Z = \frac{1}{N}\sum_{i=1}^{N} f_i \quad , \tag{2}$$

Then the normalized etalon knowledge is reduced to

$$Z_p = 1 \quad , \tag{3}$$

while the normalized knowledge of student is

$$Z_s = \frac{1}{N}\sum_{i=1}^{N} f_i \quad . \tag{4}$$

In the framework of the knowledge measurement, the following factors are important: (i) For the objectivity of the estimation the student should not be foreknown on the sequence of the asked questions. This means that the order of the records in the task (both questions and predefined answers) should be randomized. (ii) The number of questions should be large enough to achieve of the desired accuracy of evaluation.

In result the student's knowledge after the test may be represented as

$$Z_s = \frac{1}{n}\sum_{i=1}^{n} f(\xi_i) \ . \tag{5}$$

where the integer numbers ξ_j already are not the successive integer numbers of the questions in the initial collection, but the random integers distributed in [***1*** ,***N***]. With Eq.(5) the problem of the knowledge measurement may be redefined as the problem of numerical evaluation of the sum Z_s.

Mentioned above allows us to apply for such an evaluation the well-known technique of the statistical modeling (the Monte Carlo method [2]). Then the idea of evaluation takes following form: Since ξ_j are random numbers the Z_s the Eq.(5) is random also. From Eq.(5) and with the use of the central limit theorem[5] we can write the density of probability $p(Z_s)$ as follows:

$$p(Z_s) = \frac{n}{\sqrt{2\pi\sigma}}\exp\left(-\frac{\left(Z_s - \overline{f}\right)^2}{2\sigma^2} n\right), \tag{6}$$

where $\overline{f} = \frac{1}{n}\sum_{i=1}^{n} f(\xi_i)$ and σ is the standard deviation for $f(\xi_i)$. From (6) the probability for the quantity $\left|Z_s - \overline{f}\right|$ to be less than 3σ/√n is given by

$$P(\left|Z_s - \overline{f}\right| < 3\frac{\sigma}{n^{1/2}}) = \frac{1}{\sqrt{2\pi}}\int_0^3 \mathrm{e}^{-z^2/2}\, dz \approx 0.997 \tag{7}$$

One can see that for a large number of the records (questions) in the test ***n>>1,*** the student's knowledge Z_s, may be evaluated as $Z_s \approx \overline{f}$ independently on the details of the used random distribution of ξ_i , as it is written in Eq.(5). But at limited number of questions ***n,*** we have to optimize the distribution p(ξ) to minimize σ in Eq.(6a), and to improve the accuracy of the evaluation. The question arises: what distribution ***p(x)*** is good enough? To see that, we have to optimize parameter σ in Eq.(6a).

We rewrite Eq.(5) as follows

$$nZ_s = I = \sum_{i=1}^{n} f(\xi_i) = \sum_{i=1}^{n} g(\xi_i)p(\xi_i) = \overline{g}\,, \tag{8}$$

where ***g(x)=f(x)/p(x)*** and ***p(x)*** is the unknown normalized probability density ($\sum_{i=1}^{n} p_i = 1$) of the random integers ξ_i distributed in the interval ***[1, n]***, here $\overline{g}$ is the mean value of the g.

Calculating the standard deviation for random ξ_i with the use of Eq. (8) is given by

$$\sigma^2 = \overline{(g-\overline{g})^2} = \overline{g^2} - \left(\overline{g}\right)^2 = \sum_{i=1}^{n} g_i^2 p_i - \left[\sum_{i=1}^{n} g_i p_i\right]^2 = \sum_{i=1}^{n} g_i^2 p_i - I^2 \tag{9}$$

Since in (9) the quantity ***I*** is independent on the distribution ***p(x)***, we have to find such ***p(x)***, which minimizes the standard deviation σ in (9). With the use of the Cauchy-Buniakowski-Schwarz theorem [18], from (8)-(9) we obtain the following inequality

$$\sum_{i=1}^{n} g_i^2 p_i \equiv \sum_{i=1}^{n} \frac{f_i^2}{p_i} \le \sum_{i=1}^{n} \frac{f_i^2}{p_i^2} \cdot \sum_{i=1}^{n} p_i = \sum_{i=1}^{n} \frac{f_i^2}{p_i^2}\,, \tag{10}$$

where normalizing of ***p(x)*** was taken into account. The optimum of σ^2 arrives when in Eq. (10) the exact equality is achieved. Let us try to use the distribution p_i in the form

$$p_i = \frac{f_i}{\sum_{i=1}^{n} f_i} . \tag{11}$$

Substituting (11) in (10), we obtain in the left side

$$\sum_{i=1}^{n} \frac{f_i^2}{p_i} = \sum_{i=1}^{n} \frac{f_i^2}{f_i} \sum_{j=1}^{n} f_j = \left[\sum_{j=1}^{n} f_j \right]^2 = I^2 , \tag{12}$$

while the right side becomes

$$\sum_{i=1}^{n} \frac{f_i^2}{p_i^2} = \sum_{i=1}^{n} \frac{f_i^2}{f_i^2} \left[\sum_{j=1}^{n} f_j \right]^2 = \left[\sum_{j=1}^{n} f_j \right]^2 = I^2 \tag{13}$$

We observe from (11), (12) that for distribution (11) in Eq. (10) the exact equality becomes. For this case in (9) we have σ=0 , therefore now the maximal accuracy of the evaluation achieves.

However it is clear that such a choice $\boldsymbol{p_i}$ in general is unachievable, since the function $\boldsymbol{f_i}$ is unknown before the student finishes the test. Therefore, we can only recommend that to improve the test estimation, we have to choose the distribution $\boldsymbol{p_i}$ proportional to $\boldsymbol{f_i}$. In context of the knowledge estimation this means that the probability $\boldsymbol{p_i}$ to ask a i-th question has to be higher in the area, where the knowledge of the student is better.

This observation allows us to propose the following two-step adaptive algorithm. In the first step the normal exam must be made to obtain answers $\boldsymbol{f_i}$, and in the second step on the base obtained f_i, we can construct the desired distribution $\boldsymbol{p_i}$ (11), which generates new exam. Such approximation will improve the final accuracy of the evaluation. It is worth noting that the area, where the student's knowledge is poor ($\boldsymbol{f_i=0}$), already does not contribute in the total measured knowledge $\boldsymbol{Z_s}$.

Furthemore, the use of this information yields the possibility to evaluate the stability of the student's knowledge. We have to compare the student's answers received in the first step $f_i^{(1)}$ with the answers obtained in second step $f_i^{(2)}$, and then calculate the quantity S as follows

$$S = 1 - \frac{1}{n} \sum_{i=1}^{n} (f_i^{(2)} - f_i^{(1)}) . \tag{14}$$

The quantity $\boldsymbol{S}$ in (14) may be regarded as a stability of the student knowledge. If $S \approx 1$ the student has answered equally on the same questions both times, so he/she has a stable knowledge. Otherwise, if S is small the knowledge is not stable: in such a case the student's answers to the same questions were different.

The possibility to measure the stability of the student's knowledge is a very important characteristic of every evaluating system, which has to allow the estimation both quality and the level of the education obtained in an University.

3 Algorithm and Program Realization of the Online Testing System

In this section we analyze the realization of our Online Testing System (OTS) and its importance at the pedagogical process. First, we study the use of learning objects and the IMS QTI standard version 1.2.1 in our system. Then we analyze the Client-Server schema on which such OTS is based. We review the Server side, the Client side and the tools used for implementation. Finally, we show preliminary results obtained with the use of this system and discuss further work.

It is convenient to separate the program realization in two parts. The student part (Clients side) can be putted in the networks computers, while the part of code with the etalon knowledge is placed into other computer (Server side). The Server evaluates the result of exam in response of the data receiving from a student (Clients part).

3.1 The Application of Learning Objects into the OTS

In this section we analyze the meaning of learning object, how to apply this concept according the Question and Test Interoperability (QTI) standard version 1.2.1 de IMS Global to the OTS, as well as the defined migration schemas adequately to this standard.

3.1.1. What is a Learning Object?

There is a variety of definitions for learning objects that sometimes results very width, meanwhile other times are customized for the tool, system or organization that use it. The IEEE (Institute of Electrical and Electronics Engineers)'s Learning Technology Standards Committee (LTSC) defines learning object like "any entity, digital or non-digital, that can be reused or referenced by the technology supporting learning" [15]. This definition is extremely broad, and when we analyze fails to exclude any person, place, thing, or idea that has existed in any moment of the history of the universe, due any of them supports in some way learning.

Now we analyze some definitions that use the object oriented vision for computer aided instruction that turns concept confused. David Merrill uses the term "object learning" [16]. The educative software financed by the NSF (National Science Foundation) uses the term "educative software component" [10], and only accepts the Java's Applet (little Application) as learning objects [9]. The MERLOT (Multimedia Educational Resource for Learning and On-Line Teaching) Project makes reference to them as "online learning materials" [17]. Finally, Apple Learning Interchange simply refers to them as "resources" [1].

In our report we refer to learning objects as "any digital resource that can be reused to support learning" [22]. This definition includes any thing that can be delivered through the web on demand, be it large or small. Examples of smaller digital resources include digital images or photos, live or prerecorded videos, animations, or applets delivered via server, as a Java's Applet calculator [6].

3.1.2. The IMS QTI Standard Version 1.2.1 and the OTS

A proper used online testing system can result useful for the organization and users that employ it, due it makes the process faster, saves costs and generates repositories of reusable learning objects. However if the system does not adequate to any interoperability standard, the reach is local to the organization that use only. We will define a task as a composite element which integrates the question, the possible answers, the assigned time to answer, as well as the feedback or help for the respondent. The typical task is given by:

```
1. La velocidad de la luz en un bloque de vidrio con índice refracción 1.5 es:
<html> (D)2X10<sup>8</sup> m/s</hmtl>
<html> (A)3x10<sup>8</sup> m/s</html>
<html> (B)3x10<sup>8</sup> m/s</html>
<html> (C)4.5x10<sup>8</sup> m/s</html>
Todas
No se
60
Consulta la bibliografía recomendada para este curso
16. La densidad de los datos por unidad de energía Z(E), de electrones de un metal es representad
<html><div align='center'><img src='http://localhost/test/semiconductors/images/graficazea.gif' n
<html><div align='center'><img src='http://localhost/test/semiconductors/images/graficazeb.gif' n
<html><div align='center'><img src='http://localhost/test/semiconductors/images/graficazec.gif' n
<html><div align='center'><img src='http://localhost/test/semiconductors/images/graficazed.gif' n
Todas
No se
60
Consulta la bibliografía recomendada para este curso
2. El Sulfuro de Zinc es transparente en luz visible, esto implica que su brecha de energía es:
(A)>3 eV
(B)<2 eV
(C)<2.5 eV
D)<1.77 eV
Ninguna
Todas
60
Consulta la bibliografía recomendada para este curso
```

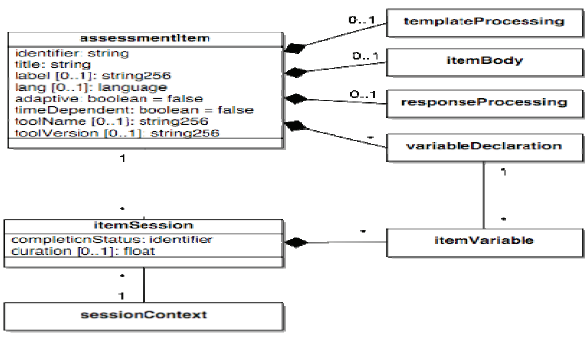
Figure 2. Segment of test tasks using proprietary standard on text-HTML

Now we discuss how we use the idea of Learning Objects in our system. The above figure represents our proprietary standard on text-HTML code for tasks management, however to being successful on knowledge economy "common standards for metadata management, learning objects and learning architectures are obligatory" [12]. Among the most important standards can be found ARIADNE (Alliance of Remote Instructional Authoring and Distribution Networks for Europe) supported by the European Union, and the American standard established by the IMS Global Consortium [8]. Keeping in mind to share our learning objects repositories, we started the migration process from our proprietary standard (see Figure 2) toward the IMS QTI standard version 1.2.1. The documentation in this standard indicates that an exam can be separated into questions or assessment items [IMS GLOBAL 2006], the rule to create these items is very simple: if the item is too large to fit the screen, then it will be necessary to create shorter items. Considering above exposed and based on the recommended UML (Unified Modeling Language) schema in Figure 3(a), first we identified required information (info.) for the assessmentItem class. Then we developed software that obtains it from the tasks declared in our proprietary standard, and automatically substitutes the information inside required sections delimited by XML tags: <assessmentItem required info. <correctResponse> required info.</correctResponse>, <outcomeDeclaration> required info. </outcomeDeclaration>, <itemBody> required info. </itemBody>,

<responseProcessing required info. />, </assessmentItem>. Figure 3(b) shows an example of the resulting XML code.

```xml
<?xml version="1.0" encoding="iso-8859-1"?>
<!--
 This example has been adapted from the PET Handbook, copyright University of Cambridge
-->
<assessmentItem xsi:schemaLocation="http://www.imsglobal.org/xsd/imsqti_v2p0 imsqti_v2p0.xsd"
        identifier="choice" title="Unattended Luggage" adaptive="false" timeDependent="false">
<responseDeclaration identifier="RESPONSE" cardinality="single" baseType="identifier">
<correctResponse>
<value>ChoiceA</value>
</correctResponse>
</responseDeclaration>
        <outcomeDeclaration identifier="SCORE" cardinality="single" baseType="integer">
<defaultValue>
        <value>0</value>
</defaultValue>
</outcomeDeclaration>
<itemBody>
<choiceInteraction responseIdentifier="RESPONSE" shuffle="false" maxChoices="1">
<prompt>La velocidad de la luz en un bloque de vidrio con índice refracción 1.5 es:</prompt>
<simpleChoice identifier="ChoiceA"> 2X10<sup>8</sup> m/s </simpleChoice>
<simpleChoice identifier="ChoiceB"> 3x10<sup>8</sup> m/s </simpleChoice>
<simpleChoice identifier="ChoiceC"> 3x10<sup>-8</sup> m/s </simpleChoice>
<simpleChoice identifier="ChoiceD"> 4.5x10<sup>8</sup> m/s </simpleChoice>
<simpleChoice identifier="ChoiceE"> Todas          </simpleChoice>
<simpleChoice identifier="ChoiceF"> Ninguna        </simpleChoice>
</choiceInteraction>
</itemBody>
<responseProcessing
        template="http://www.imsglobal.org/question/qti_v2p0/rptemplates/match_correct"/>
</assessmentItem>
```

(a) Assessment item schema



(b) XML code for single task

Figure 3. Assessment item schema (a) and the XML code for a single task (b), according the IMS Question and Test Interoperability standard version 1.2.1

3.2 Capabilities of the OTS

We have constructed our computer testing program with the following requirements. It must:

- Enable the repeatedly estimation of a basic knowledge level and a quickness of the correct answer finding.
- In case of problems (wrong answer) program has to show the correct answer, reference to the textbook, being training program.
- Show the final protocol, which contained the given questions, answers of the student, correct answers, and dynamics of his rating.
- Have a friendly and clear interface.

- To provide security by using DES algorithms to cipher communications (We will analyze the implementation of this specification in future paper).

In base of the above requirements, a Client-Server technology was developed, see Figure 4. The Client side represents the student or students and the server (the examiner). First, the client request for registration at the Server. The valid Client receives a collection of tasks containing the questions for the exam [13]. The client is asked for each task, and system stores all partial answers. When the test finished, such information is sent to the Server and processed, then the results are shown at the client side. The process of evaluation can be repeated several times in training mode and just one time in control mode.

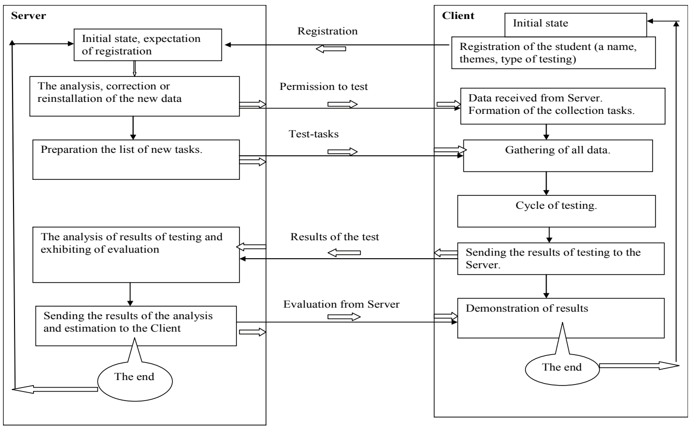


Figure. 4. Client-Server schema.

3.2.1. The Client-Server Technology

The implementation was done with the use of Object-oriented technology. We constructed the hierarchy of classes; and have code it in Java programming language [11]. The resulting technology is shown in Figure 5. At the left, is running the server program, listing the clients connected and the duration they was connected (see the AskServer section); the results obtained during examination process (see Finished Jobs); and the history of the information transferred (see historyForm) between the server and students connected.

The client is registered at the Server side, receives a collection of tasks and after that the evaluation process starts (see right side of Figure 5). The order of questions is determined by the generator of random numbers and is unpredictable. When the test is finished, the results are sent to server to be processed. Finally, the Client receives from the Server a grade, a rating and the final result of the examination.

To accumulate the results of exams, at the end of the 2005 year, we integrated MySQL Server as Data Base Management System to the OTS to avoid expensive licenses costs. Nowadays we are creating tests for several subjects to store the information in Databases and obtain the feedback from students to perform further analysis with Data Mining.

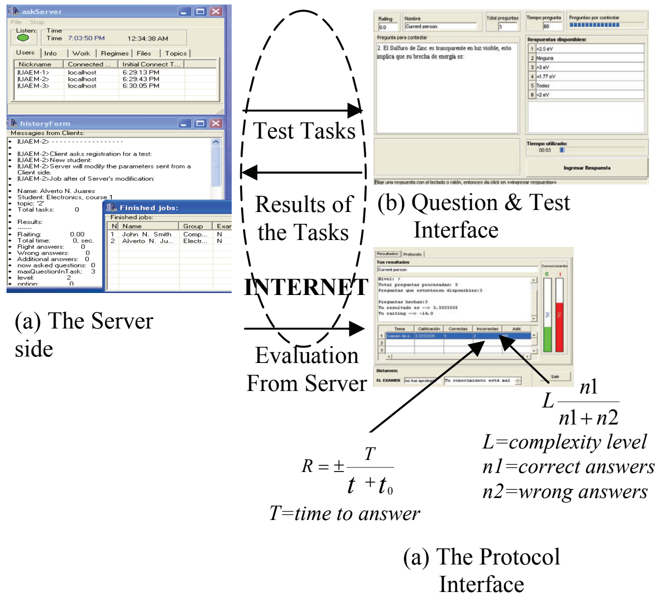


Figure 5. The server side (left) attending clients (right) [4].

3.2.2. Results of the OTS Implementation

For a simple estimation we have carried out N = 10 tests in the real tasks consisting of n = 27 questions. The average of the correct answers has appeared m = 4.5 (from 10 maximal). On the other hand, assuming the test as a gauss process, we find mathematical expectation m0=N/p, where p - probability of success in individual test, in our case p = 1/6 (6 closed answers). For the above given numbers one can find m0=4.5. The number of successes, equals 5 for this case (10 is maximum grade), is far below of the least satisfactory grade. Therefore we can declare the impossibility to obtain the satisfactory grade in absence of knowledge.

3.2.3. Further Work

To reinforce the innovative approach of our reasearch, we are going to use data mining [23] over resulting databases to determine student's learning patterns, and to avoid cheating (making frauds) on online tests. To improve the measurement of student's knowledge, we are planning to expand our system reachness from the use of question and test learning objects (assestment stage) to incoporate full learning objects that includes objective, theory, simulation and assessment. Simultaneaously, we are developing a tool to evaluate the quality of LOs since an integral approach that includes student and expert-professor (qualitative and quantitative) perspectives, as well as the use of intelligent agents and visual data mining to improve the LO's and student's learning experience [21].

4 Conclusions

The studied system can be customized for each student and carried out at any time, any where at a minimum cost. Thanks to the use of internet free of examiner; topics from different knowledge areas can be easily adapted to work within the system. The interfaces of programs (server side and client side) were designed in a user friendly way and implemented for various operating systems to avoid the platform incompatibility. System can evaluate mathematical, chemical formulas, images and even video, however performance is limited by the available Internet broad band access. The system can be used to implement surveys, to train personnel, and to assess at the successful candidates for job positions.

We can conclude that: The choice of the random order of asked questions allows estimating the level of student's knowledge at any distribution of random order of testing. At the fixed number of asked questions the accuracy of evaluation is better, if more questions are asked from the field, where the knowledge of the student is good.

The local form of the student's knowledge representation is not important, what is important only is the proportionality to the correct knowledge. For example, in a simplest case one can use the choice $f(x)$ as following: $f(x)=1$ at right answer and $f(x)=0$ otherwise. Such the method of the testing can be easily algorithmized and can be used at parallel computer evaluations of various students' groups separately in any time any where.

Acknowledgements. This work was partially supported by CONACyT granted by project 47220.

References

1. ALI. (2000). "Apple learning interchange website". In: http://ali.apple.com/
2. Bernd A. Berg (2004), "Markov Chain Monte Carlo Simulations and Their Statistical Analysis (With Web-Based Fortran Code)", World Scientific, ISBN 981-238-935-0
3. Bugai, Y., Burlak, G., Demchenko, A., Kuz'menko, N. (1997) "On Application of Computer Technology Testing In Educational Process", In: Conference Role of Universities in the Future Information Society RUFIS'97, Prague, p. 141-144.
4. Burlak, G. Hernández, J. Zamudio, A. (2005) "The Application Of Online Testing For Educational Process In Client-Server System", In: Proceedings IADIS International Conference ICWI-05, Portugal, Lisboan.
5. Box, George E.P., Hunter, William G., Hunter, J. Stuart (2002). "Estadística para Investigadores. Introducción al Diseño de Experimentos, Análisis de Datos y Construcción de Modelos", Editorial Reverté, México, D.F. p. 41.
6. Cavalli, A., Magg, S., Papagiannaki, S., Verigakis, G. (2005), "From UML models to automatic generated tests for the dotLRN e-learning platform", In: Electronic Notes in Theoretical Computer Science http://www.elsevier.com/locate/entcs, Vol. 116. p. 133-144.

7. C.P. Robert, G. Casella (2004). "Monte Carlo Statistical Methods" (second edition). New York: Springer-Verlag, 2004, ISBN 0-387-21239-6
8. Duval, E. (2004) Learning Technology Standardization: Making Sense of it All, International Journal on Computer Science and Information Systems, Vol. 1, Issue 1, p. 33-43.
9. EOE. (2000). "Educational objects economy website", In: http://www.eoe.org/eoe.htm
10. ESCOT. (2000). "Educational software components of tomorrow website", In: http://www.escot.org/
11. Javax.crypto.spec (Java2) (2005), In: http://java.sun.com/j2se/1.5.0/docs/api/javax/crypto/spec/package-summary.html, July 2005
12. Hodgins, W. (2005), "Learning Objects & Learning Standards", In: http://learnativity.com/standards.html, July 2005
13. Hunt, N., Hughes, J., Rowe, G. (2002), "Formative Automated Computer Testing (FACT)", In: British Journal of Educational Technology, Vol. 33, No 5, p. 525-535.
14. IMS Global Learning Consortium (2006) "IMS Question & Test Interoperability Specification", In: http://www.imsglobal.org/question/, July 2006
15. LTSC. (2000a). "Learning technology standards committee website". In: http://ltsc.ieee.org/
16. Merrill, M.D. (1991) et al. "Instructional transaction theory: An introduction. Educational Technology", 31(6), 7-12.
17. MERLOT. (2000). "Multimedia educational resource for learning and on-line teaching website". In: http://www.merlot.org/
18. Mitrinovic, D.S., Pecaric, J.E. And Fink, A.M. (1993). Classical and New Inequalities, Kluwer Academic Publishers, Dordrecht.
19. Questionmark Software Perception-Product Information. (2006), In: http://www.questionmark.com/us/perception/index.htm, May 2006
20. The Blackboard Learning System™ (Release 6) (2006), In: http://bb1.ctdlc.org/, May 2006.
21. Varughese, J. (2005) "Testing, Testing: IHEs are trying to ensure the success of students by testing their abilities using the latest online assessment software", In: University Business http://www.universitybusiness.com, p. 59-65.
22. Willey, D. (2000). "Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy". In D. A. Wiley (Ed.), The Instructional Use of Learning Objects: Online Version, In: http://reusability.org/read/chapters/wiley.doc, July 26, 2005
23. Witten, I., Frank, E. (2000) "Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations" Morgan Kaufmann Publishing USA p. 369.