Optimal Estimation of Penalty Value for on Line Multiple Choice Questions Using Simulation of Neural Networks and Virtual Students' Testing

Hassan M. H. Mustafa, Mohammed H. Kortam, Ibrahim H. Assaf, Ayoub Al-Hamadi, Nada M. Al-Shenawy

Abstract—This paper provides interesting findings for modeling of a challenging and critical pedagogical issue namely online learning assessment of Multiple Choice Questions (MCQs) analysis and evaluation. More precisely, in fulfillment of that issue's objective, this work suggests using a realistic Artificial Neural Network (ANN) model. That, explicitly, characterized by two learning paradigms: supervised learning (with a teacher), and autonomous (self-organized) learning. Furthermore, a computer learning assessment package used for online testing exams adopting a group of virtual 500 students. Herein, a special attention has been paid in order to search for optimal estimated penalty value. In case of multiple erroneous (incorrect) selected answers for random twenty questions submitted by any arbitrary virtual student member out of 500 virtual students. Interestingly, obtained results in case of testing two penalty values (zero & one third) shown to have bell shape close similar to Gaussian statistical distribution. Furthermore, these results have become in agreement with learning achievement results, after running of simulation program of adopted realistic ANN model considering different learning rate values.

Index Terms—Artificial neural network modeling; Computer-Aided Assessment packages; Individual differences; Multiple Choice Questions Testing; Online Self Learning; Summative/Formative Assessment; Virtual Testing.

I. INTRODUCTION

The field of learning sciences is represented by a growing community conceiving knowledge associated with educational system performance as well as assessment of technology-mediated learning processes [1]. Therefore, a recent evolutionary trend has been adopted by educationalists as well as learners due to rapid technological and social changes. So, they are facing increasingly challenges arise in this time considering modifications of educational field applications [2]. During the nineteenth of last century, educationalists have adopted Computers and Information technology in order to perform deep changes in mathematics [3][4][5]. In this context, two interesting findings have been announced by Horgan, and Aragón at [6][7]. Respectively, both findings are presented as follows."Computers are transforming the way mathematicians discover, prove and communicate ideas" [6]. And "Computers and computation have changed the entire modern world, but their effects in the fields of sciences and engineering have been especially deep" [7]. Additionally, in the same context; assessment of learning and teaching processes of a selected mathematical topic using Artificial Neural Networks (ANN²) modeling have been recently published at [1][2]. At both research papers, authors considered learning convergence (response) time, as a measuring parameter for learning process assessment. Readers are highly recommended towards more illustration, at the attached appendices shown by the end of both assigned research papers. Therein, three print screen samples have pointed to time response assessment examples in details. Additionally, in recent years several techniques have been developed aiming to reach optimality of large-scale assessments of student achievement in mathematics [8]. Conversely, to time response parameter considered at [1][2], this piece of research introduces, students' achievements for learning measurement associated with obtained outcomes of Multiple Choice Questions (MCQ²) assessment. More specifically, this piece of research herein, inspired by a strong belief that interdisciplinary combining of Artificial Neural Networks (ANN²) modeling. With observed challenging, and critical phenomenal educational issues for their investigation, analysis, and evaluation. In addition to on line virtual students' testing via packaged computer program for the assessment of any student's achievement (selected answers). Accordingly, firstly realistic ANN model has been adopted in performing simulation for an interesting critical pedagogical issue. That concerned with estimation of formative and summative assessment of online learning processes outcomes. Briefly, the basic definitions of both distinct educational assessment types have been presented at[9][10]. In the context of (ANN), this work considers a realistic (ANN) model which explicitly characterized by two learning modes: supervised learning (with a teacher), and autonomous (self-organized) learning.[11][12].
Secondly, by running implemented simulation software package on computer, it results in interesting analysis and evaluation of summative assessment procedures. In harmony with procedures of formative assessment following after published findings at [13]. The simulated software package presents both types of on-line self-assessment utilizing MCQ approach while submitting exams at classrooms. Via adopted assessment package, individual difference phenomenon has been observed among samples of virtual students during estimation of penalty value. This phenomenon of individual students' differences could be realistically considered as classified after Myers-Briggs Type Indicator (MBTI) (Briggs Myers, 1962) [14]. That MBTI is based on Jung’s theory of psychological types (Jung, 1923)[15][16]. When students are involved in criteria and goal setting, self evaluation becomes obviously a logical step in the learning process. In this case, students become meta-cognitive and are more aware of their personal strengths and weaknesses. Referring to Black & Wiliam, "When students are required to think about their own learning, articulate what they understand, and what they still need to learn, achievement improves." [17]. In classrooms, student's ability to access his or her measurement of own performance, is an interesting and important educational issue. That could be beneficial in building up lifelong skills that will be used in work and interpersonal relationships. In the educational context, self-assessment is classified as either summative or formative. There is a distinct difference between both [18]. Firstly, summative assessments given periodically to determine at a particular point in time what students know and do not know. However secondly, in case of formative assessment, part of the instruction process informs both teachers and students about student understanding at a point when timely adjustments can be made. The obtained estimated optimal of penalty value for self-assessment students' considered students' individual differences. Specifically, the presented case study of learning assessment herein based on randomly selected group of 20 multiple choice questions. Each group considered as an exam submitted to 500 virtual students. Noting that any of submitted questions includes a set of four multiple choice solutions. At Figure 3, a simplified macro-level flowchart describing algorithmic steps considered on-line self assessment. Interestingly, obtained self assessment results have been by statistically presented by graphs. Their described statistical distributions (for given two penalty values) shown to have bell shaped curves which similarly near to Gaussian (normal) distribution as shown at Figure 6 & Figure 7. These two illustrative figures have been well supported by results concerned with creativity as published at [19]. Interestingly, the obtained statistical analysis results of penalty value are shown to have well analogy. With learning model using (ANN) associated to individual differences at two distinct learning rate values ($\eta$ =0.1 & $\eta$ =0.5)[1] (Referring to Figure 8 in below). It is worthy to note that penalty for four questions by MCQ exam suggested at [20], has been considered to perform verification of presented penalty value by this paper results.

The rest of this paper organized in four sections as follows. At next section, revising of assessment basic concepts are briefly introduced. Assessment by simulation of ANN model with analogy to supervised/unsupervised learning modes is presented at section III. Furthermore, description of suggested Computer-Aided Assessment (CAA) package via simulation of virtual students also given at the same section. At section IV the simulation results are introduced in details. Finally, some conclusion remarks are given at section V.

II. REVISITING OF ASSESSMENT BASIC CONCEPTS

In this section, a brief presentation for educational assessment phenomenon basics have been introduced. It is well known that assessment issue has a broad and rapidly growing field, with a strong theoretical and empirical base [22]. However, an assessment expert couldn’t be found who is capable to employ sound practices to guide any teaching process. That for obtaining more systematical investigation towards planning as well as implementation for programmed of observed learning assessment phenomenon. The basic concepts needed to more systematic investigation of suggested assessment analysis and evaluation are presented at [21][22]. The next three subsections (A, B, and C) introduce a brief revising for formative assessment, summative assessment, and student's self-assessment respectively:

A. Formative Assessment

Formative assessment is also known as Assessment for Learning that is defined as the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there”. It do not result in an evaluation. Information about what a student knows, understands and is able to do is used by both the teacher and the learner to determine where learners are in their learning and how to achieve learning goals. In more details, this kind of assessment refers to the gathering of information or data about student learning during a course or program that is used to guide
improvements in teaching and learning. Furthermore, Formative assessment activities are usually low-stakes or no-stakes; they do not contribute substantially to the final evaluation or grade of the student or may not even be assessed at the individual student level. For example, posing a question in class and asking for a show of hands in support of different response options would be a formative assessment at the class level. Observing how many students responded incorrectly would be used to guide further teaching [22]. Therefore, this paper adopts that model capable for learning assessment via online testing for a virtual group of 500 students. These students have been subjected to twenty Multiple Choice Questions (MCQs), which concerned with some computer science curriculum. Herein, attention paid to search for optimal estimated penalty value in case of erroneous (incorrect) selected answers. These selected answers have been assigned freely (with randomized probability value) by any arbitrary virtual student member out of 500. Furthermore, By using suggested ANN model, a fairly unbiased estimated penalty values have been obtained after measuring simulated outcomes of proceeded MCQ's virtual testing sessions. During examination sessions, fixed number of MCQs have been submitted to the student. Interestingly, examined students have provided with their ability to skip (freely at his request) all provided answers for any submitted question.

B. Summative Assessment

Conversely, to the above concept presenting formative assessment at subsection A., summative assessment is known as Assessment of Learning result in an evaluation of student achievement - for example, allocation to a level or standard or allocation of a letter or numerical grade, which might later appear in a report [21]. The goal of summative assessment is to evaluate student learning at the end of an instructional unit by comparing it against some standard or benchmark. Furthermore, unlike formative assessments' activities which usually have low-or no-stakes; often summative assessments have high stakes, which means that they have a high point value. They are usually included as examples: a midterm exam, a final project, a paper, and a senior recital. Interestingly, information from summative assessments can be used formatively when students or faculty use it to guide their efforts and activities in subsequent courses [23]. Therein, at that reference, exploration for the extent to which assessment information can be used for both summative and formative purposes, without the use for one purpose endangering the effectiveness of use for the other.

c. Student self-assessment [24]

In this Professional Learning module, the term 'student self-assessment' is used as an umbrella term which encompasses: student self-assessment; student self-evaluation; and student self-regulation or self-monitoring. In other words, the focus is given towards the ability of the students to: understand both learning intentions and success criteria, use these criteria to judge what they have learnt and what they still need to learn, reflect on the learning process to ascertain how they learn best, act on feedback received from their teacher and their peers, set learning targets based on what they still need to learn, and manage the organization of their learning. Student self-assessment encourages students to take responsibility for their own learning. It incorporates self-monitoring, self-assessment and self-evaluation.

III. MODELING OF LEARNING ASSESSMENT

A. Generalized ANN Model

Essentially, assessment functional model given as diagrammatic view for an interactive educational process, at Figure 1. in the above. It illustrates generalized ANN block diagram which representing analogous simulation of both two online learning assessment paradigms. More precisely, presented model is well qualified in performing realistic simulation of either summative or formative assessment functions as follows. By more details, three vectors representing analogously inputs to the neural network as learning model at that Figure 1. are provided by environmental stimuli(unsupervised learning).The correction signal for the case of learning with a teacher is given by responses outputs of the model will be evaluated by either the environmental conditions (unsupervised learning) or by the teacher. Finally, the tutor plays a role in improving the input data (stimulating learning pattern), by reducing noise and redundancy of model pattern input by giving correction of submitted answer. For package model presented in this work, correction of submitted answer is inherently stored according to relation between Computer Aided Learning (CAL), and -Aided Assessment (CAA) packages. [25], additionally, presented automated (CAA) package herein could provide both the student and the lecturer with feedback on accessed learned course material [26].
B. Mathematical Formulation Of the ANN Model

The presented model given at Figure 2, simulates generally two diverse learning ANN modes that analogous to supervised/unsupervised assessment. It presents realistically both paradigms: by interactive learning / teaching process, as well as other self organized (autonomous) learning. By some details, firstly is concerned with classical (supervised by tutor) learning observed at our classrooms (face to face tutoring). Accordingly, this paradigm proceeds interactively via bidirectional communication process between tutor and his/her learner(s). However, secondly other learning paradigm performs self-organize (autonomously unsupervised) tutoring process. The mathematical formulation for the suggested ANN model is given as follows.
\[ \tilde{e}(n) = \tilde{y}(n) - \tilde{d}(n) \]  

Where

- \( \tilde{e}(n) \): Error correcting signal controlling adaptively the learning process (Vector).
- \( \tilde{y}(n) \): Output signal of the model (Student's Answer).
- \( \tilde{d}(n) \): Objectively desired numeric value(s).

Generally as a learning process parameter (Correct answer vector).

Referring to above figure 1, following four equations are deduced:

\[ V_l(n)=X_l(n) W_{lj}^T(n) \]  
\[ y_l(n)=\phi(V_l(n))= (1-e^{-\lambda y_l(n)}) / (1+e^{-\lambda y_l(n)}) \]  
\[ e_l(n)=|d_l(n)-y_l(n)| \]  
\[ W_{lj}(n+1)=W_{lj}(n)+\Delta W_{lj}(n) \]

Where: \( X \ldots \) input vector, \( W \ldots \) weight vector, \( \phi \ldots \) is the activation function, \( y \ldots \) is the output, \( e \ldots \) the error value, and \( d \ldots \) is the desired output.

Noting that \( \Delta W_{lj}(n) \) the dynamical change of weight vector value.

Above four equations are commonly applied for both learning paradigms: supervised (interactive learning with a tutor), and unsupervised (learning though students' self-study).

The dynamical changes of weight vector value specifically for supervised phase is given by equation:

\[ \Delta W_{lj}(n)=\eta e_l(n) x_l(n) \]  

Where \( \eta \) is the learning rate value during learning process for both ANN learning modes.

However, for unsupervised paradigm, dynamical change of weight vector value is given by equation:

\[ \Delta W_{lj}(n)=\eta y_l(n) x_l(n) \]  

Noting that \( e_l(n) \) in (6) is substituted by \( y_l(n) \) at any arbitrary time instant \( n \) during learning process. In the assessment context, seeking and interpreting evidence for use by learners and their teachers after some consecutive time instants consecutive \( n \), reaches correct (desired) answer. This type of assessment called (Assessment for Learning) or equivalently Formative assessment. Unlike this type of assessment, summative assessment is known as Assessment of Learning result in an evaluation of student achievement by completely ending of learning process. The statistical distribution of obtained simulation results for that ANN model has been presented at Figure 9. This Figure illustrates individual differences at two distinct learning rate values (\( \eta =0.1 \& \eta =0.5 \)).

C. Computer-Aided Assessment (CAA) package

The adopted package for assessment of both types considered for optimal estimation of penalty value in case of multiple erroneous (incorrect) selected answers by any arbitrary (random) virtual student member out of 500 virtual students. These selected answers have been assigned freely (with randomized probability value) by any virtual student for MCQ-Exam (20 questions). Furthermore, examined students have provided with their ability to skip (freely at his request) all provided answers for any submitted question. Specifically, attention herein paid to search for the possibility to validate Multiple Choice Questions as introduced at [19]. The macro-level flowchart for the adopted (CAA) package is given at next subsection.

D. Package Macro-Level Flowchart

In this section, main module of adopted (CAA) package is presented as inspired by diagrammatic view for interactive educational process (at Figure 1). The macro-level flowchart that describes algorithm of adopted (CAA) package at Figure 3. It gives a well description for basic algorithmic steps for both self assessment approaches (Summative and Formative) for MCQ-Exam (20 questions). Accordingly, after running of suggested computer package, that results in performing fairly and unbiased penalty values' estimation. Therefore, running of that package, results in verification of suggested penalty value (1/3) as given at [19].
Figure 3. A simplified macro-level flowchart briefly describes algorithmic steps considering on-line self assessment.

E. Print Screen Samples
This subsection includes three illustrative distinct figures for print screen sample cases of print screen. They are obtained after online assessment by interaction of virtual randomly selected student out of 500 while answering a set of Multiple Choice Questions. Referring to three Figures{(3)&(4)&(5)}, they consist of three print screen snapshots illustrating both summative as well as formative on line assessment of any arbitrary virtual student.
Figure 4. The observed student's summative final on-line assessment resulting by the end of answering all 20 questions' exam case if the student has chosen correct answer.

Figure 5. A Print Screen that illustrates the case if an arbitrary virtual student has selected his/her answer correctly. Hence, that student will be informed that his/her selected answer is O.K., by observing a marked (green) correct sign. Accordingly, this process belongs to formative on-line self-assessment approach.

Figure 6. A Print Screen that illustrates formative on-line self-assessment when wrong answer has been selected by arbitrary virtual student. Hence, he/she will get penalty value of (-1/3), and a marked wrong (red sign) could be observed.

IV. SIMULATION RESULTS

This section introduces obtained simulation results in two forms (tabulated and graphical) at subsections A. &B, respectively. Additionally, at subsection C., graphical simulation results of obtained statistical distribution
considering samples of 1000 virtual students' achievements, is presented for ANN model presented at section II in the above.

A. Tabulated Results

<table>
<thead>
<tr>
<th>Penalty Value</th>
<th>Students' average Assessment Value (M)</th>
<th>Variance $\sigma$</th>
<th>Standard deviation $\sqrt{\sigma}$</th>
<th>Coefficient of variation $\rho = \sqrt{\sigma} / M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5.004</td>
<td>3.652</td>
<td>1.911</td>
<td>0.382</td>
</tr>
<tr>
<td>- 1/3</td>
<td>-0.158</td>
<td>6.325</td>
<td>2.515</td>
<td>2.515</td>
</tr>
</tbody>
</table>

# TABLE II. ASSESSEMENT RESULTS FOR TWO PENALTY VALUES (ZERO & ONE THIRD) WITH SKIPPING

<table>
<thead>
<tr>
<th>Penalty Value</th>
<th>Students' average Assessment Value (M)</th>
<th>Variance $\sigma$</th>
<th>Standard deviation $\sqrt{\sigma}$</th>
<th>Coefficient of variation $\rho = \sqrt{\sigma} / M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4.054</td>
<td>3.199</td>
<td>1.7886</td>
<td>0.442</td>
</tr>
<tr>
<td>- 1/3</td>
<td>0.125</td>
<td>5.268</td>
<td>2.295</td>
<td>18.361</td>
</tr>
</tbody>
</table>

B. Graphical Results

Figure 7. Illustrates the statistical distribution for two penalty values (0 & -1/3) with availability for skipping the answer any arbitrary question.
V. CONCLUSION

The interesting concluded remarks are given as follows:

1) This work presents well justification for optimal penalty value (-1/3) as suggested therein at[20]. That is verified via statistical distribution of virtual students’ answers after MCQ testing as given at two Figures (6&7).

2) The misleading penalty value during MCQ testing will result in false indication about actual simulation results of learning rate values.

3) Obtained results herein seemed to be reasonably valuable for illustration of more promising systematical investigations for other MSQ examination instructions considering selection of one correct answer out from two, three, or five answers as presented at adopted reference [20].

4) The presented learning assessment (summative/ formative) for any arbitrary student either carried out via guidance of his/her teacher or by self learning. Both educational processes are shown to be respectively; well analogous to the performance of ANN model either supervised (see equation (6)) or unsupervised learning (see equation (7)). That illustrated well at subsections (A. &B.) in section III.
5) Finally, in future, extension of this work recommended investigating the summative and/or formative learning assessment phenomenon in more elaborated approach using realistic ANN simulation and modeling.

REFERENCES


[20] An Artificial Neural Networks Examination, June 2004 which includes Multiple Choice Questions with availability to select only one answer out of 2, 3, 4, or 5 that suggested to get penalty value of -1 & -1/2 & -1/3 & -1/4 respectively; in case of choice one erroneous answer http://aas.oru.se/~tdt/ann/tenitor/exam2004_2.pdf


[26] Question Mark for Assessment Management, Create, Deliver, and Deliver. Available online at: https://www.questionmark.com/us/Pages/default.aspx
