Our era is characterized by information overload. Low skills are a demand in the world of work, such as salespeople in brick-and-mortar stores, customer reception, fast-food restaurant clerks, call center operators, order dispatchers, order packers, etc. The required educational background shrinks to the requirement for adaptability and few “basic skills”: text comprehension, basic communication in one or two foreign languages, some math, science and technology concepts, a good dose of familiarity with digital as well as interpersonal and social skills (OECD, 2001). Assessment in secondary education has not remained unaffected; the holistic nature, originality, and innovation are incompatible with everyday teaching practice. With the scientific tools of statistics and neural networks, the divergence between formal assessment and diagnostic and feedback assessment has been quantified. Finally, the evolution of behavioral learning at a public high school level during the academic year 2021-2022 has been studied.

**Keywords:** formative evaluation, neural networks, regression analysis, summative evaluation.

I. INTRODUCTION

Education is a dynamic process that continues to evolve and is constantly shaped. Similarly, the evaluation process is also evolving and is shaped, and we could say that it obeys the rule expressed by Heraclitus: “Everything flows.”

Assessment in our current era could be said to have three basic forms: assessment in the classroom, assessment in national exams, and assessment of increasingly difficult subjects, i.e., in mathematical competitions either at the national level or at the global level (Dylan Wiliam, 2007; Jan de Lange, 2007).

Could we recognize some connection between these three forms?

Let’s try to imagine these three forms in 3D space. First of all, we would have our basic level, which would be the classroom assessment, then a parallel level, which would be the national exam assessment, and a third parallel level, which would be the assessments of increased difficulty.

The relationship between these three levels is linear. However, if we project the two levels onto the first, we will see that their intersections are very small, as success in classroom assessment does not guarantee success in national exams, and success in national exams does not guarantee success in competitions of increased difficulty.

Does the assessment have a specific structure, or is it a process that is applied depending on the teacher or the curriculum each time?

As our educational system does not examine the students holistically, but essentially only the skills acquired by the students in the application of the algorithms required to solve the exercises. So, the evaluation, simulating the demand of our time, which is nothing but control over inferior skills, is divided into stages and procedures (Peña-López, 2012).

The definition of these procedures obviously simplifies the evaluation, but at the same time also creates questions regarding these procedures.

1) What will be the Framework of the assessment?
2) What are the functions that this Framework will define?
3) How will the assessment be measured and
4) whether this measurement will be valid?

Although in national exams there is a specific Framework, the classroom assessment setup is completely different. (Fig.1). Each teacher adjusts the evaluation according to the level of his/her students. Students from different schools in the same grade are evaluated in a completely different manner; therefore, the result is that the grades are almost virtual since, on the one hand, students are examined in subjects that they have already been taught many times or if the level is raised trying to solve unknown exercises (Suurtamm et al., 2016).

---

**Published Online:** November 21, 2022
**ISBN:** 2736-4534
**DOI:** 10.24018/ejedu.2022.3.6.498

D. Varveris*
International Hellenic University of Greece campus Kavala, Greece.
(e-mail: jimvarveris1968@gmail.com)

V. Tsiantos
International Hellenic University of Greece campus Kavala, Greece.
(e-mail: tsianto@physics.ihu.gr)

V. Saltas
International Hellenic University of Greece campus Kavala, Greece.
(e-mail: saltas70@gmail.com)

*Corresponding Author
A. Validity of the Evaluation

By 2022, the so-called topic bank will be implemented for the second time in our country. The first time the topic bank was implemented in 2014, and this lasted only for one year and then stopped due to a change of education policy by the new Ministry of Education administration.

In other words, we see the effort of external evaluation that will provide objectivity in terms of the general evaluation of the students. It can be noted that the subject bank is applied only to the upper high school classes (grades 10-11) and not to the lower high school classes (grades 7-9). So, the evaluation gets its validity since the student is not evaluated only by the teacher who teaches him but also by a common base of subjects (Fig. 2) (Hamilton et al., 2007).

Moving to a non-subjective evaluation with graded exercises leads to holistic thinking as opposed to assessing the student's weaknesses in the daily schedule, where the student is assessed on problems of mechanic and algorithmic procedures (Palm et al., 2011).

B. Measurement in Evaluation

The building blocks of the forms of evaluations are unchanged, (Fig. 2). The foundation of the edifice is the structured nature which is independent of its trappings and members of evaluation (Kaarstein, 2014).

With regards to the functionality and context of the evaluation, there is a large gradation of multiple tests and processes on which the students must be evaluated. The evaluation is done in writing, but in recent years a combination with digital means has been employed. It has also been observed that the written exams or digital exams have not presented differences in evaluation, so there is independence in the means of evaluation (Wang et al., 2007).

As can be seen in the bibliography, the examination using printed material did not contribute much to the improvement process (Wiliam, 2007). Therefore, in the formal assessment that has been done so far by using the implementation of the digital transition, the learners, instead of solving the exercises in notebooks, are assessed on online platforms. This is an intermediate stage toward diagnostic assessment (Wiliam, 2007).

The conditions and building blocks of the arithmetic approach of the evaluation and the measurement itself are often not identified according to the behaviorist school (Thorndike, 1922; Hull, 1943; Skinner, 1954; Gagne, 1965).

(Fig. 2). Such as the family environment of the students, their economic situation and the special characteristics of each student are not examined. The attempt to measure the result affects the result itself, i.e., in quantum physics, where the arithmetic result of the calculations includes the error resulting from the measurement itself. Therefore, the question arises as to whether learning can be considered as having an increasing correlation between transmitter and receiver, i.e., a quantitative increase in stimulus and responses (Thorndike, 1922; Hull, 1943; Skinner, 1954; Gagne, 1965).

C. Summative and Formative Evaluation

Evaluation without feedback, unfortunately, does not contribute to learning. For the development and improvement of students, the evaluation must be a tool that signals the weaknesses and learning gaps of each student, so evaluation must be a teaching tool (Delandshere, 2002; Broadfoot & Black, 2004; Gipps, 1994; Delandshere, 2002; Shepard, 2000).

The summative assessment is a tool for categorizing students, i.e., generally speaking, there are five qualities of students, bad, moderate, good, very good, and excellent. So, evaluation is an information tool that helps the teacher to put each student into the category he/she belongs to. Unfortunately, the traditional evaluation does not show the student what his/her weaknesses are. Furthermore, it does not show the student the ways to improve himself/herself. It is useful to just label the student, where it acts as a self-fulfilling prophecy, and the student does not change his/her behavior, but he or she fits into the above-mentioned qualities.

In contrast to that, formative assessment helps the teacher with a diagnosis of the student's gaps and, at the same time, confirms the same using feedback if these gaps have been overcome or still exist so that the student does not stagnate in one situation but moves onto a better situation.

If we assume that tutoring takes place at a time interval of t0, studying at t1, and assessment at t2, we observe the sequential and linearity Fig. 3 of the three phases of learning.

Space and time are the reason why the three phases are asynchronous (Graue, 1993).

Let's imagine a different Spatio-temporal Fig. 4 model of learning:

D. Evaluation Variables

The Greek Ministry of Education has created assessment variables by revising Presidential Decree 126/2016 in the article on the assessment of the performance of high school students, which is replaced as follows:
"Article 3-Evaluation procedure:
The procedure for assessing the student's performance is defined as follows:
The following criteria are taken into account for the evaluation of the student's performance during the four quarters:

1) The student's overall participation in learning instruction (the questions he/she asks, the answers he/she gives, his/her contribution to the study of a topic in class, his/her cooperation with classmates, the diligence in carrying out the tasks assigned to him/her), from which the teacher gets an idea of the student's knowledge, understanding of concepts and phenomena, problem-solving skills, communication skills, critical thinking, creativity, etc.

2) The work carried out by the pupil in the context of the daily learning process at school or home, individually or in groups.

3) Creative compositional work, individual or group work, interdisciplinary work, individual or group work.

4) Quarterly assessment tests (hourly written tests or assignments and submission/presentation of individual or group compositional or interdisciplinary creative work or use of the features and stages of the inverted classroom model).

5) Short written tests (tests)...

In this paper, the correlation between the variables, i.e., short written tests (test) and quarterly assessment tests, have been studied by regression analysis.

E. Research Aims and Questions

The present research aimed to study short written and quarterly assessment tests through regression analysis, making it possible to evaluate and assess student performance.

1) Student Assessment and evaluation using quarterly and other short tests. How did students perform during the short and quarterly evaluations?

2) Regression between the variables. Does the value of regression reflect the causality relationship between our variables?

3) What were the primary factors that affected their performance?

4) What were some observable changes, and what did they indicate?

II. METHODS

A. Data Sample

In this research, short written and quarterly assessment tests were studied using the regression analysis technique to study the correlation between them. The research aimed to assess and evaluate performance using created assessment variables. The quarterly tests contained assignments, group and individual presentations, and other submissions. Short written tests contained short, relevant tests that students had taken.

B. Data Collection Methods

Since these tests had been taken and their marks recorded, there were no specific data collection methods. Therefore, various reliable data sources were consulted to get this information. Verified data from one public school 7th grade class were obtained to effectively get marks and specific grades for the targeted tests. The obtained data were grouped depending on factors such as the time taken for easy and effective studying. This also ensured that the results we got were not biased. Secondary data is in refined form; hence collection process was effective, time-saving, and easy. The collection methods focused primarily on grades from short written and quarterly assessment tests.

C. Neural Networks Approach

1) Overview

The neural network approach helped us know if the regression value reflects the minimal causality between the variables.

Many think neural networks are only used in machine learning. However, it can also be used in data analysis research and in many other areas. It is very effective in turning raw and unstructured data into viable and valuable information. Companies that deal with a large amount of data use it to extract and obtain relevant information from their vast raw data. Neural network models have an input, hidden, and output layer. The input layer is the first layer in a neural network model. It has input neurons that allow raw materials to be brought into the model for processing purposes. The hidden layer exists between the output and input layers. Most non-linear transformations are conducted in this layer. The output layer is the final part of the model that produces the final prediction.

In this case, with the help of a neural network, it was possible to extract relevant data from collected records. Our network had an input layer with two factors and a number of units which was 21. The number of units in the hidden layers was 7, and the activation function SoftMax. The output layer gave out dependable variables, the number of units, the error function, the activation function, and the recalling method. The first process was to prepare the datasets from all the obtained records. Once the preparation phase was over, it was integrated and loaded into the model. Evaluations were done. The patterns that emerged were noted. Hidden patterns were easily recognized over time and classified. Each decision was classified depending on its weight. After training the network, the value of mse was obtained.

D. Validity and Reliability

To ensure there was validity and reliability in the research, expert advice and assistance were sought from various evaluation specialists and sources. The specialists also rechecked and validated data from there cords sources to ensure the final results were relevant and unbiased. In case of errors, the data sources were contacted for clarifications, and afterward, necessary corrections were carried out after approval. Most of these test evaluations were conducted by a common subject teacher, further validating the information we got. The given evaluation variables ensured the process was reliable. The obtained secondary data were analyzed to come up with relevant results Fig. 5. Linear regression was used to assess dependable evaluation variables.
III. RESULTS

A. Regression Analysis for the Variables Test Fractals and Test Line and Plane

First, it is examined whether the degree of test fractals relative to the degree of test line and plane gives us a good regression.

<table>
<thead>
<tr>
<th>TABLE I: MODEL SUMMARY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

a. Predictors: (Constant), Test line and plane.

In the Model Summary table (Table I), we are interested in the Adjusted square term, which is 0.520 and expresses the percentage of total variability present in the data that is explained by the regression.

It has a low value that minimizes the possibility of interpretation between the two tests since the acceptable value is above 0.75.

We then do a t-Test (Table II) to identify which b is not 0.

<table>
<thead>
<tr>
<th>TABLE II: COEFFICIENTS*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>(Constant)</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

a. Dependent Variable: Test fractals.

\[-y = b_0 + b_1x_1 \quad (1)\]

where:
- Dependent=Constant B + coefficient of independent B independent variable. B_i
- In today’s case, y=0.760*B_i
- If p=Sig.<0.05 x is statistically significant.

B. Regression Analysis for the Variables First-Period Test and Test Line and Plane

In the Model Summary table (Table III), the interest is in the Adjusted R-square term, which is 0.659 expressing the percentage of total variability present in the data and interpreted by the regression. So, the value of the Adjusted R square term is a better value than in Sub heading A and this means there is an improvement between the test and the test of quarter A.

<table>
<thead>
<tr>
<th>TABLE III: MODEL SUMMARY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

a. Predictors: (Constant), Test line and plane.

With a t-Test (Table IV), we determine which b is not 0.

<table>
<thead>
<tr>
<th>TABLE IV: COEFFICIENTS*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>(Constant)</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

a. Dependent Variable: First period test.

In the regression the coefficients are:

\[b_0=0 \text{ and } b_1=0.830.\]

So, \(y=0.830*x_1.\)

C. Regression Analysis for the Variables First-Period Test and Test Fractals

In Table V, the interest in the Adjusted R square term is 0.662. It is noted that it is almost identical to the regression for the variables first-period test and test line and plane (Sub Heading B).

<table>
<thead>
<tr>
<th>TABLE V: MODEL SUMMARY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

a. Predictors: (Constant), Test fractals.

We make a t-Test (Table VI).

<table>
<thead>
<tr>
<th>TABLE VI: COEFFICIENTS*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>(Constant)</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

a. Dependent Variable: First period test.

In the regression, the coefficients are:

\[b_0=5.310 \text{ and } b_1=0.809\]

So:

\[y = 5.310 + 0.809x_1 \quad (2)\]

D. Causality Between the Variables Test Fractals, Test Line and Plane, and First Period Test

The regression between the variables studied is found to be at low values, so the question arises: does the low value of the regression reflect the minimal causality relationship between our variables?
The answer will be given by the neural networks (Table VII) from the value of the error in mse. (Table VIII).

### Table VII: Network Information

<table>
<thead>
<tr>
<th>Input Layer</th>
<th>Factors</th>
<th>Test line and plane</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Units</td>
<td>2</td>
<td>21</td>
</tr>
<tr>
<td>Hidden Layer</td>
<td>Number of Units</td>
<td>Test fractals</td>
</tr>
<tr>
<td>Dependent Variables</td>
<td>1</td>
<td>First period test</td>
</tr>
<tr>
<td>Output Layer</td>
<td>Number of Units</td>
<td>1</td>
</tr>
<tr>
<td>Rescaling Method for Scale</td>
<td>Standardized</td>
<td></td>
</tr>
<tr>
<td>Activation Function</td>
<td>Identity</td>
<td></td>
</tr>
<tr>
<td>Error Function</td>
<td>Sum of Squares</td>
<td></td>
</tr>
</tbody>
</table>

a. Determined by the testing data criterion: The “best” number of hidden units is the one that yields the smallest error in the testing data.

We observe the high correlation of our variables expressing high causality, which is nothing but the study of the students themselves (Table VIII).

### Table VIII: Model Summary

<table>
<thead>
<tr>
<th>Training</th>
<th>Sum of Squares Error</th>
<th>0.953</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative Error</td>
<td>0.127</td>
<td></td>
</tr>
<tr>
<td>Training Time</td>
<td>0.00:00:01</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Testing</th>
<th>Sum of Squares Error</th>
<th>0.001*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative Error</td>
<td>0.037</td>
<td></td>
</tr>
</tbody>
</table>

a. The number of hidden units is determined by the testing data criterion: The “best” number of hidden units is the one that yields the smallest error in the testing data.

The value of mse after training the network is 0.037 and expresses both high correlation and predictive ability.

### IV. Discussion and Conclusions

The regression analysis for the variables tests fractals and test line and plane gave a very low correlation. We could interpret this as follows: Students did not take their grades in the 1. test seriously, so they did not enhance and change their study mode for the 2. test. The summative assessment did not change their behavior and especially did not diagnose their weaknesses so that they could improve.

The regression analysis for the variables, such as the first-period test and test line and plane, gave a higher correlation but was not satisfactory. Students paid more attention to the quarter A test as participation in the grade was more important. Weaknesses and deficiencies remain at high levels.

The regression analysis for the variables first-period test and test fractals also gave almost the same correlation as the regression analysis for the variables of the first-period test and test line and plane. Confirming the greater attention to the test in quarter A but also the inability to identify and change students’ behavior by improving them.

The causality between the variables tests fractals, test line, and plane, and first-period test, as we found, is very high with almost zero error. The high causality is based on the intuitive variable personal study, demonstrating the importance of personal study in learning and that knowledge acquisition is mainly a solitary process.

### APPENDIX

**A. Test on fractions**

1) **Theme 1**

Calculate the algebraic expressions:

\[ A = 4 + 3 \cdot \frac{1}{3} - \left( \frac{3}{2} - \frac{1}{4} \right) \cdot \frac{3}{2} \]

and

\[ B = 2^5 - 29 \cdot 1^{29} + (6^2 - 4 \cdot 9) \cdot 29 + 5^2 + (2^2 + 1^{20}) \]

a) Show that A=9/2 and B=8.

b) Calculate Γ=4:A-2:B.

2) **Theme 2**

Three friends shared 600 euros. The first one got 2/3 of the amount, the second one got 1/4 of the amount, and the third one got the rest.

a) How many euros did each person get?

b) What fractional part of the amount did the third person get?

3) **Theme 3**

A farmer sold to three traders 2/3, 1/3 and 1/10 of his wheat production. Calculate:

a) What part of his production remained unsold??

b) If he had 2000kg left, what would be his total production?

c) How many kilos of wheat were 2/5 of his production?

**B. Test straight-line and plane**

1) **Theme 1**

a) Give the definition of a Straight-line segment?

b) Give the definition of semi-straight-line, straight-line?
C. First period exams

1) Subject 1st

Give the definitions and draw them on a plane:
Straight-line segment, semi-straight line, straight line, semi-straight lines objects.

2) Subject 2nd

Calculate the expression:

\[ A = \left( \frac{11}{3} - \frac{3}{7} \right) \div \frac{4}{5} \]

\[ \left( \frac{1}{6} + \frac{2}{7} \right) \div \frac{3}{5} \]

3) Subject 3rd

Let the angle be 45°.

a) Find 1/3 of its Supplementary and 1/3 of its complementary.

b) To be designed so that:

1) Be adjacent Angles.
2) Don't be adjacent Angles.

CONFLICT OF INTEREST

Authors declare that they do not have any conflict of interest.

REFERENCES


