analytics

MDPI

Article

Can Oral Grades Predict Final Examination Scores? Case Study in a Higher Education Military Academy

Antonios Andreatos [1,*] and Apostolos Leros [2,†]

[1] Division of Computer Engineering and Information Science, Hellenic Air Force Academy, Dekeleia, 13671 Acharnes, Greece

[2] General Department, National and Kapodistrian University of Athens, 15772 Athens, Greece; lerosapostolos@gmail.com

* Correspondence: antonios.andreatos@hafa.haf.gr

† Prof. Apostolos Leros has been retired from the National and Kapodistrian University of Athens.

Abstract: This paper investigates the correlation between oral grades and final written examination grades in a higher education military academy. A quantitative, correlational methodology utilizing linear regression analysis is employed. The data consist of undergraduate telecommunications and electronics engineering students' grades in two courses offered during the fourth year of studies, and spans six academic years. Course One covers period 2017–2022, while Course Two, period 1 spans 2014–2018 and period 2 spans 2019–2022. In Course One oral grades are obtained by means of a midterm exam. In Course Two period 1, 30% of the oral grade comes from homework assignments and lab exercises, while the remaining 70% comes from a midterm exam. In Course Two period 2, oral grades are the result of various alternative assessment activities. In all cases, the final grade results from a traditional written examination given at the end of the semester. Correlation and predictive models between oral and final grades were examined. The results of the analysis demonstrated that, (a) under certain conditions, oral grades based more or less on midterm exams can be good predictors of final examination scores; (b) oral grades obtained through alternative assessment activities cannot predict final examination scores.

Keywords: assessment; oral grades; final exam; correlation; prediction; linear regression; engineering course; alternative assessment

check for updates

Citation: Andreatos, A.; Leros, A. Can Oral Grades Predict Final Examination Scores? Case Study in a Higher Education Military Academy. *Analytics* **2023**, *2*, 836–852. https://doi.org/10.3390/analytics2040044

Academic Editors: Vivekanandan Suresh Kumar, Kinshuk and Jong-Min Kim

Received: 30 December 2022
Revised: 24 August 2023
Accepted: 7 September 2023
Published: 2 November 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

1.1. Objective of the Study and Research Questions

Academic achievement is a complex process of student engagement in educational procedures and varies greatly from student to student for the same course, as well as from course to course for the same student, as faculty members can assure. Assessment of academic achievement represents a measurement of students' knowledge, skills, and abilities.

The aim of this research is to investigate the possible correlation of student performance as expressed in the oral grades and the written final examination grades in a Greek military academy (henceforth "the Academy"), located in Attica, Athens' metropolitan area.

The Academy offers, among other specializations, a degree in telecommunications and electronics engineering (TEE). This program is equivalent to a bachelor's degree in electrical engineering [1]. In contrast to universities, engineering classes have a small number of students.

The data were collected from two technology-related undergraduate 400-level courses:

- Course One, Fall 2017, Fall 2019, and Fall 2021 semesters;
- Course Two, period 1: Fall 2014, Fall 2015, and Fall 2017 semesters;
- Course Two, period 2: Fall 2019, and Fall 2021 semesters.

The two courses were taught by different teachers to the same students at the same period of time. Two datasets were derived per semester per course: oral grades and final exam scores. In both courses, the final written exam typically consisted of a set of open-ended questions and problems, similar to those included in the textbooks.

1.1.1. Alternative Assessment

Alternative assessment is a method of evaluation that measures a student's level of proficiency in a subject as opposed to the student's level of knowledge (typically measured by exams). The concept of alternative assessment is to allow students to acquire and demonstrate knowledge and skills by performing tasks. Alternative assessment is also called performance testing or authentic assessment, because it is deeply rooted in one's ability to do something by leveraging newly-gained knowledge. As part of the assessment, the students have to perform meaningful tasks that reflect a clear understanding of the teaching and learning objectives. Since Course Two is mostly a practical subject, it makes sense to adopt alternative assessment [2].

1.1.2. Course One

Course One is a theoretical technological undergraduate 400-level course offered by the Division of Electronics, Electric Power and Telecommunications to the TEE cadets of the Academy during their seventh semester of studies. The syllabus includes optical fibers, optical sources (LED and laser), photodetectors (photodiodes), optical couplers, optical filters, optical amplifiers, etc. This course assumes two prerequisite courses: applied electromagnetics and solid state electronic devices, as well as a strong mathematical background.

The recommended textbook is S.O. Kasap's, *Optoelectronics and Photonics*, 2nd edition, Pearson, 2013. There is also a supplementary recommended textbook: *Optical Fiber Communications*, 3rd ed., by G. Keiser, McGraw Hill, 2000. In addition, instructor's notes and lecture slides are used. Oral grades result from a midterm exam. The educational policy of the course remained unchanged for the duration of this study.

1.1.3. Course Two

Course Two is a mostly practical technological undergraduate 400-level course offered by the Division of Computer Engineering and Information Science to the telecommunications and electronics engineering cadets during their seventh semester.

Course Two covers an introduction and application and transport layers. The textbook used is J. F. Kurose and K. W. Ross' *Computer Networking: A Top-Down Approach*, 7th edition.

For many years, oral grades were derived from homework assignments and a midterm exam, based on the textbook problems or similar problems from the bibliography, with the midterm exam counting for 70% of the oral grade. From Fall 2019 onward, oral grades in Course Two have been derived from a set of alternative academic activities offered throughout the semester, instead of exams or tests. These activities typically include educational scenarios, lab exercises, simulations, etc. [1,3]. The set of activities changes every year, to adapt to technological developments, as well as to the students' interests.

The reasons behind this decision were the following:

- Course Two is a practical subject. Lab exercises and educational scenarios are invaluable for practical courses and enjoyable for the students. On the other hand, they facilitate and deepen learning;
- Different students have different cognitive abilities; using a variety of alternative activities caters for more learning styles;
- There were students who could not perform well in written exams due to stress, although they were trying hard in the classroom. Assessment in Course Two should have pluralism, in order to be fair.

After this change in the oral assessment policy, the teacher of Course Two noticed a significant difference between oral grades and final examination scores; there were students who performed well on alternative academic activities but poorly on the final exam and vice versa. This observation was one of the motivations for this research.

1.2. Motivation and Objective of This Work

One field of pedagogical research is investigating the relationship of oral grades to final examination performance. If the two datasets are correlated, then oral grades could serve as a predictor of final grades. If this is the case, then low oral grades could be used as an early warning of student failure in the course.

It is important to identify at-risk students at an early stage of their academic career, so that strategies to support them can be put into action [4–6]. Early prediction of student course failure is important in our academy, and gives the potential to provide students with timely feedback and support.

Since there is not enough research literature regarding the use of oral scores as a predictor of final exam scores [7], this study represents a contribution towards this direction.

A second objective of this study was to investigate the association between different oral assessment methods and the level of academic performance in the final written examination. Three oral assessment methods were considered: exam-based oral assessment (Course One); oral assessment based on lab exercises and homework assignments, and a midterm exam (Course Two period 1); and assessment based exclusively on alternative activities (lab exercises and educational scenarios, Course Two period 2 [3]).

This comparative, quantitative, and correlational research used regression analysis to construct predictive models for the Course One and Course Two final grades. Hence, the research questions are as follows:

Research question 1: Can the oral grades based on homework assignments and a midterm exam be used as a predictor of the final examination scores?

Research question 2: Can the oral grades based exclusively on alternative assessment activities be used as a predictor of the final examination scores?

Consequently, the purpose of this research was to examine if there is any relationship between the oral grades and final examination grades for undergraduate telecommunications and electronics engineering cadets in computer technology courses, in three different cases: when the oral grades resulted from exams, when the oral grades resulted from homework assignments and exams, and when the oral grades resulted from alternative assessment activities.

2. Literature Review

2.1. Assessing Academic Achievement

Assessment is a necessary and continuous educational task that gives valuable information about a students' progress, as well as valuable feedback to instructors about their teaching practice. Various types of assessment methods measuring student academic performance are available to educators.

Academic achievement depends on various factors including the students' personality, self-efficacy [8,9], intrinsic and extrinsic motivation [10], background, effort, age, and cognitive skills, as well as the teacher's emotional and academic support, and parental and social support [11,12]. Teacher engagement and student motivation are main areas of research. Students who perceive their instructors as more supportive achieve better academic performance [13–20].

Grades are the most common means of assessing students' academic performance [21]. Grading provides datasets of students' academic achievements [22], facilitating storage, ordering, comparison, etc. Students who complete the required weekly reading achieve higher scores in course assessments. A multiple-test strategy was found to increase overall test scores and promote learning [23,24]. Students' ability to score well in exams depends on several factors other than knowledge, including psychological and psycho-social factors, as well as personal skills. A student's motivation, background, self confidence, stress, and time management skills affect their exam performance [23,25–28].

Lack of proper guidance, a lack of or inappropriate recommendations, last minute changes, and unfair means and policies severely affect students' exam performance [29]. Stress is another important issue; students experience high levels of stress, which often leads to anxiety before and during exams and ultimately affects their academic achievement [30]. Therefore, the effectiveness of tests is disputed [31]. Grade-based assessment can disorient educational goals and may encourage students to cheat; many researchers criticize these forms of exams as unfavorable and not constructive for students [32].

On the other hand, academic achievement has been identified as a high source of anxiety within university students. Sometimes, the relationship between anxiety and academic performance is indirect, because of the valuation of auto efficacy, meaning that people with a low perception of auto-efficacy generate high levels of anxiety [33]. A significant inverse relationship between test anxiety and course grades, in both undergraduate and graduate students, has been reported in the literature [34–36].

2.2. Alternative Forms of Assessment

Fortunately, educators have a number of tools at their disposal to help students [10]. Alternative forms of assessment encourage behavior that results in increased learning, information retention, and student achievement [2,27,37].

During alternative assessment activities, the instructor acts more as a facilitator than as an evaluator, thus the students are not stressed. For this reason, from Fall 2019 onward, oral grading in Course Two was based on alternative assessment activities rather than tests, quizzes, or exams [37].

2.3. Related Work

Various data mining techniques such as classification and regression [38] have been applied to build student performance prediction models. Classification is used when the outcome variables are categorical (or discrete), while regression is used when the outcome variables are numerical (or continuous) [39].

Classification is the most commonly applied data mining technique in higher education [40]. The most popular classification algorithms used to predict student performance are naive Bayes, K-nearest neighbor, and decision tree [39]. Since the data in this research were numerical, we will next focus on regression-based studies.

Aissaoui et al. [39] used multiple linear regression to determine the relationship between a dependent variable (students' performance) and many independent variables. Their dataset included 32 demographic attributes and 395 records. The aim of their work was to select the most important among the attributes, in order to build a multiple regression linear model that enabled the prediction of the students' final grade. After determining the most important variables resulting from each method, they used those selected variables to build multiple linear regression models. The most popular variables among the seven models produced were mother's education, mother's job, father's education, extra educational support, and going out with friends. Then they used R^2 to measure the correlation between the actual outcome values and the values predicted by each model. By comparing the performances of the models, they found that the best model was the one created using the "MARS" method [39].

Alabbad et al. [41] investigated the reliability of medical student logbook data for assessing student's performance and predicting outcomes of an objective standardized clinical exam (OSCE) and a multiple-choice questions (MCQ) exam during surgery rotation. Univariate linear regression analysis was used to evaluate the associations of the number of clinical encounters and the number of clinical tutors with OSCE and MCQ scores. No correlation between the volume of self-reported clinical encounters and exam scores was found [41].

Davison and Dustova [7] used linear regression analysis to investigate predictive relationships between standard examinations (standard true/false and multiple-choice questions) and practical examinations (hands on system administration tasks) for undergraduate students for two courses of a computer technology program. The first course was a 200-level course focusing on systems administration. The second course was a 300-level computer technology course focusing on infrastructure services.

For the 200-level course, it was found that oral grades did not significantly predict the written final exam scores. For the 300-level course, the linear regression analysis resulted in a statistically significant predictive model. However, since in both cases R^2 was close to 0, they concluded that the resultant models were not a good fit, as they suffered from a high unexplained variance [7].

Cui et al. report that for a science course, a predictive model, which was built on data from one semester, was able to identify about 70% of students who failed the course and 70% of students who passed the course in another semester, with only LMS data extracted from the first four weeks [6].

3. Methodology and Design

In this research, where the data were numerical, a linear regression analysis was used to produce predictive models between oral grades and final exam scores. This design type also allowed for hypothesis testing. The methodology selection was driven by the research questions. For the above reasons, a quantitative methodology was selected, utilizing a correlational study.

3.1. Research Hypotheses

- Null Hypothesis: The oral grades do not significantly predict the written final exam grades.
- Alternative Hypothesis: The oral grades do significantly predict the written final exam scores.

3.2. Variables

In this study, the oral grades were used as the independent variable and the final examination scores as the dependent variable. Correlation between variables does not necessarily imply causality.

3.3. Environment and Control

Both courses met once a week for three academic hours (120 min), face-to-face, for 15 weeks during the fall semester (September to mid-January). The midterm written examinations (where applicable) lasted three academic hours (2 h and 15 min). The final written examinations were set at the end of each semester (January and June) and lasted approximately 3 h; they were given on paper and were manually graded by the instructors. Both midterm and final exams were supervised by an instructor or staff members. Grades ranged from 0 to 100. In general, final exams are more difficult than midterm exams, because they examine all (or a large part) of the syllabus, may combine issues from various chapters, and typically contain more complex problems.

3.4. Data Collection

Both courses were offered in the fourth year of studies. After 2015, the TEE specialization was offered biennially, hence both courses were offered in Fall 2017, 2019, and 2021.

In this study, aggregated data from academic years 2014–2015, 2015–2016, 2017–2018, 2019–2020, and 2021–2022 were analyzed. In all cases, the whole class population participated (100%). For each case we considered, we took the most recent semesters. The data encompassed three sets, each consisting of oral grades and final grades.

- Dataset 1 consists of 33 pairs of exam-based oral and final examination scores and came from Course One Fall 2017, 2019, and 2021.
- In dataset 2, the oral grades resulted from homework assignments, lab exercises, and a midterm exam, accounting for 70% of the oral grade; this came from Course Two Fall 2014, 2015, and 2017, and consisted of 37 samples;

- In dataset 3, the oral grades resulted exclusively from alternative assignment activities; it came from Course Two Fall 2019 and 2021, and consisted of 24 samples. In this case, there was no third semester, so the sample was smaller than for the other two cases.

Consequently, the purpose of this research was to investigate possible relationships between oral grades and final grades in these three datasets. The grades were analyzed in terms of correlations and score prediction. Next, three predictive models were created: one for Course One, and two for Course Two.

3.5. Demographics

Sixty-one students participated in this study. The students were in their fourth year of studies, about 21 years old, 87% males and 13% females.

The teachers of both courses are senior full-time professors, men, with over 25 years of experience, having taught these courses for at least 16 years.

3.6. Data Analysis

The grades were analyzed in terms of statistics and correlations using the Calc software package (part of LibreOffice version 6), and Excel's Data Analysis ToolPak (part of MS Office 2021). The results were verified in Matlab, Octave, and PSPP, a free program for statistical analysis of sampled data (https://www.gnu.org/software/pspp, accessed on 10 March 2023). The resultant predictive models were derived from the analysis.

The metrics used to evaluate the models were the correlation R, R squared (R^2), the standard error, the ANOVAs, the *p*-value, the root mean squared error (RMSE), etc. [7,39].

The correlation coefficient R, or Pearson's r, is a measure of the strength and direction of the linear relationship between two variables; it is defined as the covariance of the variables divided by the product of their standard deviations and represents the relationship between two variables.

When R is close to zero, there is no relationship between the variables. When it is close to 1 or −1, there is a strong relationship between the two variables. The square of the correlation coefficient R^2 is often used instead of R. In an ideal model, R^2 should be close to 1. An R^2 close to 0 indicates no relationship.

The mean squared error (MSE) measures the average of the squares of the errors between the real and estimated final grades. The square root of the MSE, called RMSE, is often used instead. The lower the MSE and RMSE, the better the forecast.

The standard error is the average distance of the observed values from the regression line.

Analysis of variance (ANOVA) determines the influence of the independent variable on the dependent variable in a regression study.

Significance F is the *p*-value associated with the F-test of overall significance. This test determines whether a model does a better job explaining the dependent variable's variability than a model with no independent variable, and informs us whether or not the regression model is statistically significant. Here, we assumed a confidence interval of 95%. Thus, if the *p*-value was less than 0.05, there would be a statistically significant association between the oral and final grades.

An outlier is a point far away from the regression trend line. Such cases are possible and may be due to various factors, including stress, poor planning, negligence, indifference, psychological factors, or even external causes. The presence of even one outlier in a small sample (such as ours) can dramatically deteriorate or even nullify a model.

4. Results

From our experience, we anticipated the oral activities statistics (maximum, minimum, and mean) would be higher than the final exam statistics, because the problems in the final exam are typically more complicated than those of the midterm exam.

Using linear regression, models predicting the final examination score (dependent variable y) using the oral grades as a predictor (independent variable x) were constructed for each semester. The models had the form of a linear equation, where α is the coefficient and β is the constant (or intercept) term:

$$y = \alpha \cdot \chi + \beta \tag{1}$$

4.1. Course One

In Course One, the oral grades result from a midterm exam. Next, the collective results for the three most recent semesters are presented (Fall 2017, Fall 2019, and Fall 2021).

The dataset consisted of 33 pairs of oral–final grades. The resulting model had the following characteristics (Table 1):

Table 1. Aggregate statistics for Course One (N = 33).

Statistics	Oral		Final
Maximum	93		95
Average	61.76		59.33
Minimum	40		15
Spread	53		80
Standard deviation	15.47		17.74
Variance	239.46		314.71
MSE		150.65197	
RMSE		12.27404	

The average score for the midterm exam (oral grade) was 61.76%, while the average score for the final exam was 59.33%, i.e., very close. The standard deviation for the midterm exam was 15.47%, while the standard deviation for the final exam was 17.74%, i.e., comparable. Moreover, the maximum grades were very close, but the variances presented some difference.

The regression equation is presented below (Equation (2)).

$$y = 0.82772 \cdot \chi + 8.21556 \tag{2}$$

The predictive model for Course One is presented below in Figure 1. Notice the existence of outliers.

Figure 2 describes statistically the produced linear regression model for Course One.

The regression statistics table provides statistical measures of how well the model fits the data. We can see that R^2 = 0.521295, which agrees with Figure 1.

The R-squared value of 0.521295 indicates that our model accounted for about 52.1% of the dependent variable's variance.

The standard error of the regression indicates the typical size of the residuals. This statistic shows how wrong the regression model was on average. We want lower values, because this signifies that the distances between the data points and the fitted values are smaller. Here, the standard error was 12.663785, which is rather large.

In the ANOVA table, df means degrees of freedom. SS is the sum of squares (4971.515017 in this case; see second line), and MS is the mean of squares produced as SS/df, where df is 31 here.

F is the significance, which is very good in this case, meaning that our test result was statistically significant, so the model was valid.

Table 2 summarizes the produced linear regression model for Course One, obtained using the Data Analysis ToolPak in Excel.

The regression equation was statistically significant (2.108099×10^{-6}). However, R^2 was equal to 0.506, rather moderate. The standard error of the estimate was also moderate (12.66). These facts indicate that the produced model for Course One was significant but had a moderate accuracy.

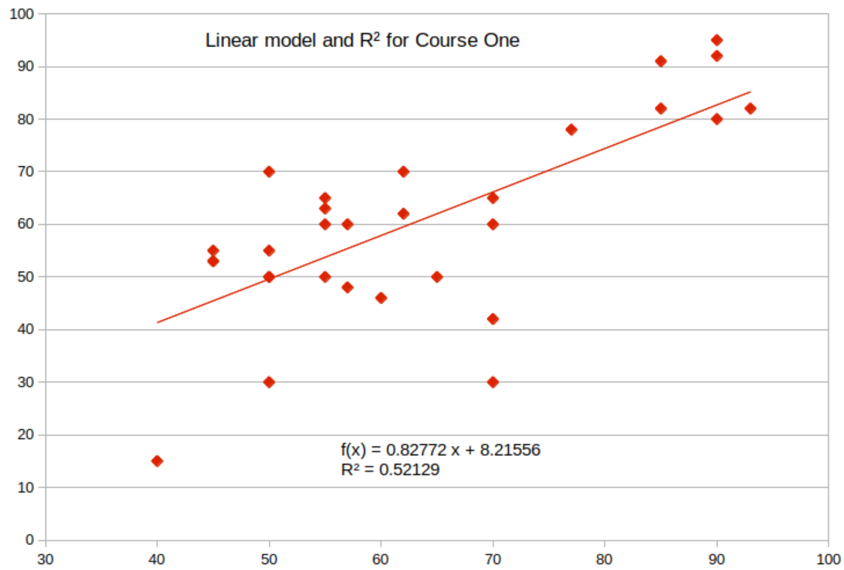

Figure 1. Linear model and R^2 for Course One.

Course One

Regression Statistics	
Multiple R	0.722007
R Square	0.521295
Adjusted R Square	0.505853
Standard Error	12.663785
Observations	33

ANOVA

Anova	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	5413.818316	5413.818316	33.757993	0.0000021081
Residual	31	4971.515017	160.371452		
Total	32	10385.333333			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	8.215563	9.069973	0.905798	0.372030	−10.28276915	26.713896
X Variable 1	0.827717	0.142460	5.810163	0.000002	0.537167	1.118266

Figure 2. Course One linear regression model statistics.

Table 2. Course One linear regression model description.

Observations	33
R	0.722007
R Square	0.505853
Sum of Squares (SS)	4971.515017
Mean Square (MS)	160.371452
Standard Error	12.663785
Significance	2.108099×10^{-6}

4.2. Course Two

In Course Two, we consider two different cases, according to the way oral grades are derived:

1. Period 2014–2017, including three semesters: Fall 2014, Fall 2015, and Fall 2017. A series of assignments are offered throughout the semester, contributing 30% to the oral grade. The remaining 70% of the oral grade comes from a midterm exam.
2. Period 2017–2019, including two semesters: Fall 2019 and Fall 2021; the oral grades are derived from a series of alternative assessment activities.

4.2.1. Period 2014–2017

The oral grades and final exam statistics for Course Two are presented in Table 3. The maximum, average, and minimum of oral grades were superior to those of the final exam, as expected.

Table 3. Oral and final grades statistics for Course Two 2014–2017 (N = 37).

Statistics	Oral	Final
Maximum	100	98
Mean	82.29	75.36
Minimum	38.5	44.5
Spread	61.5	53.5
Standard Deviation	16.43	14.92
Variance	126.17	135.66
MSE	89.307945	
RMSE	9.450288	

The above results are considered normal, according to our educational policy. In addition, the oral and final scores are comparable across all criteria (maximum, mean, minimum, standard deviation, etc.).

The regression equation (e.g., predictive model) for Course Two is presented below (Equation (3) and Figure 3).

$$y = 0.606132 \cdot \chi + 28.267864 \tag{3}$$

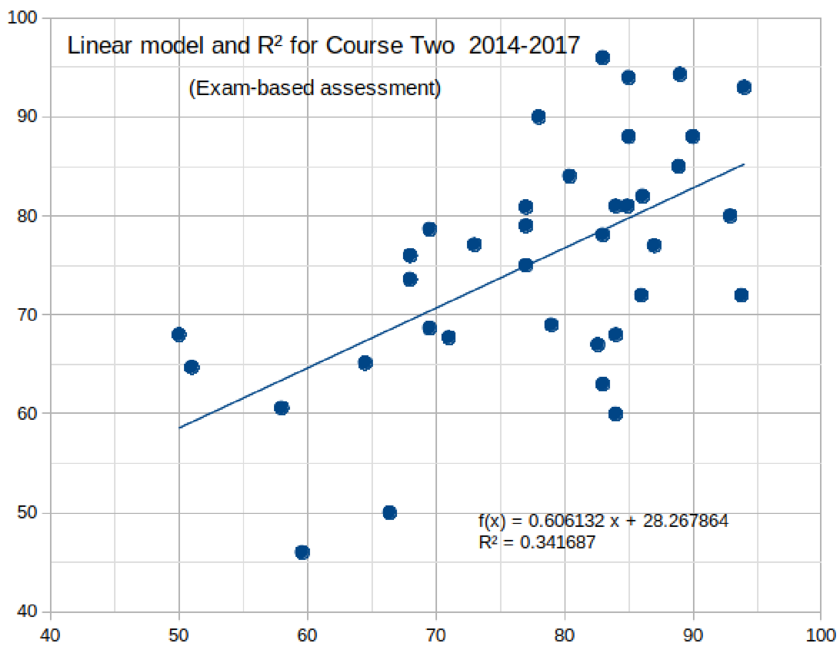

Figure 3. Linear model and R^2 for Course Two 2014–2017.

Figure 4 describes statistically the produced linear regression model.

Table 4 summarizes the produced linear regression model for Course Two, period 2014–2017, obtained by the Data Analysis ToolPak of Excel.

The regression equation was statistically significant (0.00014528). However, R^2 was equal to 0.342, rather small. The standard error of the estimate was not bad (9.7). These facts indicate that the produced model for Course Two was significant but had a medium accuracy.

Course Two 2014-2017

Regression Statistics	
Multiple R	0.584540
R Square	0.341687
Adjusted R Square	0.322878
Standard Error	9.716546
Observations	37

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1715.093454	1715.093454	18.166197	0.00014528
Residual	35	3304.393978	94.411257		
Total	36	5019.487432			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	28.267864	11.195903	2.524840	0.016261	5.538972	50.996756
X Variable 1	0.606132	0.142212	4.262182	0.000145	0.317427	0.894837

Figure 4. Course Two 2014–2017 model statistics.

Table 4. Course Two 2014–2017 linear regression model description.

Regression Statistics	
Observations	37
R	0.584540
R Square	0.341687
Sum of Squares (SS)	3304.393978
Mean Square (MS)	94.411257
Standard Error	9.716546
Significance	0.00014528

4.2.2. Period 2019–2021

The oral grades of Course Two Fall 2019 and Fall 2021 resulted exclusively from alternative assessment activities including lab exercises, educational scenarios [3], even the attendance of a free MOOC [42].

Oral grades and final exam statistics are presented in Table 5. The maximum and average of the oral grades were superior to those of the final exam, as expected. However, the spread, standard deviation, and variance of the oral grades were superior to those of the final grades, while the oral minimum was lower than the final minimum. This result is in contrast with Course One and Course Two, period 2014–2017, where the oral grades were derived from a midterm exam by 100% or 70% respectively. This means that the alternative assessment made a difference.

Table 5. Oral and final grade statistics for Course Two 2019 and 2021 (N = 24).

Statistics	**Oral**	**Final**
Maximum	100	98
Average	82.29	75.36
Minimum	38.5	44.5
Spread	61.5	53.5
Standard deviation	16.43	14.92
Variance	325.89	227.22
MSE	203.476560	
RMSE	14.264521	

The regression equation (e.g., predictive model) for Course Two Fall 2019 and Fall 2021 is presented below (Equation (4) and Figure 5).

$$y = 0.269935 \cdot \chi + 54.396116 \tag{4}$$

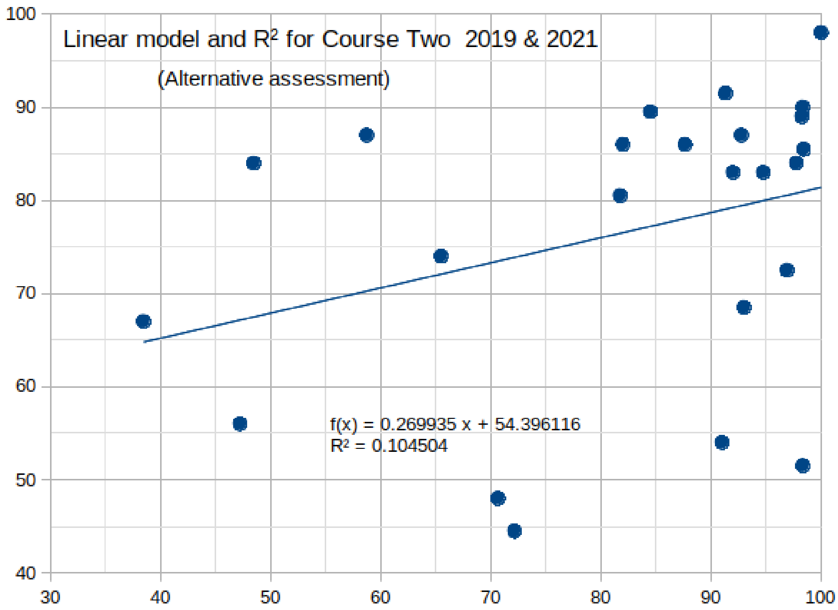

Figure 5. Linear model and R^2 for Course Two 2019 and 2021.

Figure 6 summarizes the produced linear regression model, obtained using the Data Analysis ToolPak in Excel.

Course Two 2019-2021

Regression Statistics	
Multiple R	0.323271
R Square	0.104504
Adjusted R Square	0.063800
Standard Error	14.898806
Observations	24

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	569.895898	569.895898	2.567394	0.12334941
Residual	22	4883.437435	221.974429		
Total	23	5453.333333			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	54.396116	14.227864	3.823210	0.000927	24.889332	83.902900
X Variable 1	0.269935	0.168466	1.602309	0.123349	−0.079442	0.619312

Figure 6. Course Two 2019–2021 model statistics.

Table 6 describes statistically the produced linear regression model for Course Two, period 2019–2021.

Table 6. Course Two 2019–2021 linear regression model description.

Regression Statistics	
Observations	24
R	0.323271
R Square	0.104504
Sum of Squares (SS)	4883.437435
Mean Square (MS)	221.974429
Standard Error	14.898806
Significance	0.1233494056

The generated model was statistically insignificant and exhibited poor characteristics (e.g., a small R square and large standard error). This result was expected because, as we have already mentioned, the observation that triggered this research was that there were students with a low overall performance who excelled in the alternative assessment activities, but also students with a good overall performance who performed poorly in the alternative assessment activities. The rule of thumb in such cases is the existence of points in the upper left and lower right quadrants of the regression graph(in this case, Figure 5).

4.3. Impact of Results on Hypotheses

The models produced for Course One and Course Two 2014–2017 were valid and, to some degree, accurate; consequently, they could be used to predict final grades. Therefore, the null hypothesis could be rejected for these courses; and the alternative hypothesis holds. Thus, exam-based oral grades do significantly predict written final exam scores.

On the other hand, the model produced for Course Two 2019 & 2021 was statistically insignificant and could not be used to predict final grades. Therefore, the null hypothesis holds in this case: oral grades based on alternative assessment activities do not predict written final exam scores.

5. Discussion

The oral grades in Course Two 2014–2017 are based primarily on the midterm exam (70%); this was the reason behind the good characteristics of the Course Two Fall 2017 model (since we found that exam-based oral grades can predict final exam grades).

The oral grades of Course Two Fall 2019 and Fall 2021 resulted from alternative assessment activities instead of midterm exams; as a result, the generated models were invalid, in the sense that they could not predict the final exam grades. Course Two Fall 2019 & 2021 presented almost no correlation between the oral and final grades ($R^2 = 0.1045$).

The Course One model showed that, under certain conditions, it was possible to predict final exam scores from oral scores. In this case study, these prerequisites seemed to be:

1. Similar methods of deriving oral and final grades (that is, exams).
2. Consistent educational policy (same teacher, same book, etc.).
3. Consistent student behavior. When students lose their motivation or have a weak background or when something goes wrong and they cannot perform regularly, outliers occur which deteriorate or even cancel the model.

Over time, with the aggregation of additional data, and provided that the other critical factors of the educational process remain stable, the models are expected to improve and stabilize; therefore, it is imperative to maintain historical data. When important factors change across semesters, new models must be constructed. If we have a lot of observations, it is possible, as well as acceptable, to remove extreme samples, in order to improve the model. A statistical estimator that is often "tweaked" to obtain the best possible model is the MSE (mean squared error); the smaller the MSE, the closer we are to finding the line of best fit [43]. For example, we can “tweak” the Course One model by removing the samples which generated the largest MSE between the prediction and final grade, to obtain a better model. The improved predictive model and regression equation for Course One are presented below (Figure 7).

Figure 8 describes statistically the improved linear regression model for Course One, obtained after removing the four worst samples, and Table 7 presents the regression statistics.

The produced model is significantly improved in terms of R^2, standard error, significance, etc., and hence, more accurate. This being the case, low oral grades can be used as an early warning of student failure in the final exam.

On the other hand, the failure of the model produced for Course Two period 1 is attributed to the fact that the alternative assessment activities used to produce the oral grades were different in nature and skills from the problems used in the final exam.

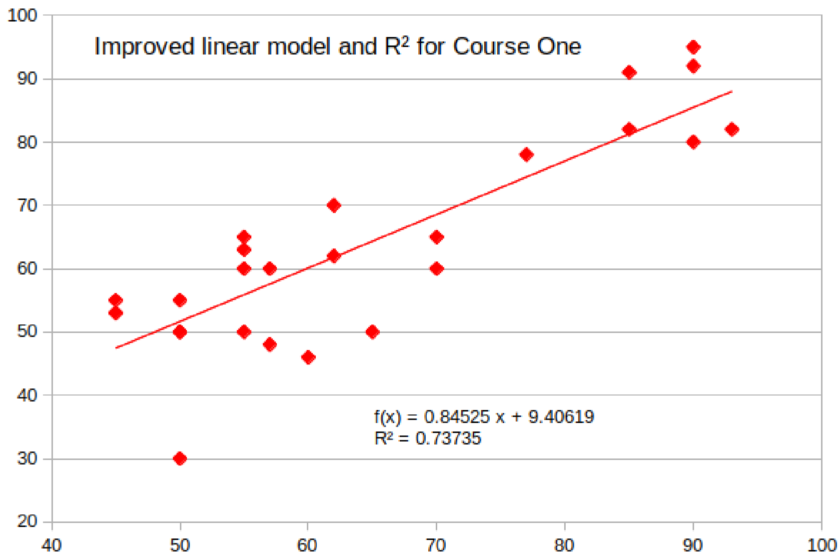

Figure 7. Improved linear model and R^2 for Course One.

Course One (improved)

Regression Statistics	
Multiple R	0.858692
R Square	0.737353
Adjusted R Square	0.727625
Standard Error	8.206199
Observations	29

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	5104.463643	5104.463643	75.799443	0.0000000025
Residual	27	1818.226013	67.341704		
Total	28	6922.689655			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	9.406193	6.241658	1.507002	0.143422	−3.400630	22.213017
X Variable 1	0.845255	0.097086	8.706288	0.000000	0.646052	1.044458

Figure 8. Course One improved linear regression model statistics.

Table 7. Course One linear regression improved model description.

Regression Statistics	
Observations	29
R	0.858692
R Square	0.737353
SS	1818.226012
MS	67.341704
Standard Error	8.206199
Significance	2.539482×10^{-9}

The model produced for Course Two period 1 is different from that of period 2. The difference between the two cases was attributed to the change in the educational policy concerning oral assessment.

The results presented here are a case study for engineering classes of the Academy, which typically have few students. In such cases it is difficult to design accurate prediction models, because the unpredictable performance of even one student can affect a large percentage of the class statistics and may invalidate the model. We expect larger samples would produce more reliable models.

This is a pioneer study concerning the Higher Education Military Academies of Greece. Its value lies in the fact that, in courses where oral grades are based on exams, the generated models per course could be used primarily as prediction tools, allowing the educators and dean to identify students at risk of failing the exam early and intervene in order to help these students improve their academic performance. The students themselves could perceive any individual deficiency in the course in time, in order to avoid failure. In this case, early awareness may bring about less accumulated stress, because the students would have more time to study consciously to avoid failure. In courses with prerequisites or in colleges with specific restrictions, where students are permanently expelled if they fail several times in a course or miss the academic year due to inadequate performance, a prediction tool would be particularly useful.

In addition, this study demonstrates that alternative assessment helps students who perform poorly in exams to learn practically. This was confirmed in an interview with one such student after his graduation [44].

For Course Two Fall 2019 and Fall 2021, the lack of correlation between oral and final exam grades was attributed to the following reasons:

1. Learning activities that contributed to the oral grades tested different skill sets from the final exams. Therefore, students' ability to execute lab exercises and homework assignments was not a good predictor of their ability to solve the problems in the final exam.
2. The students' performance changed under the different conditions that applied in the alternative activities and the final written exam and depended to a large extent on their learning style; some learning styles are favored in alternative assessment activities, while other learning styles are favored in written exams [3,44,45].

Different Courses, Different Models

Our findings also indicate that different courses produce different models; this is reasonable, since different courses can vary a great deal in terms of course contents, learning objectives, educational policy, methods of evaluation and assessment, etc. Several changes in critical parameters, such as different instructors, teaching methods, books, etc., will greatly affect the model credibility. Each course tends to have unique characteristics that affect its predictive model [6]. Moreover, the difficulty levels of different courses vary considerably, so we should not expect models from a difficult course to apply to a relatively easy course and vice versa. Our results indicate that a general model cannot address the complexity of all courses, because learning objectives, content difficulties, educational policy, activities and assessments, exam difficulty, etc., vary greatly. The use of a general model would unavoidably compromise the model accuracy in predicting student course failure [6].

6. Conclusions

This research study explored the relationships between oral grades and final examination grades for two undergraduate 400-level engineering courses, offered in a Hellenic military academy during the period 2014–2022.

A quantitative, correlational approach utilizing linear regression analysis to describe possible relationships between oral and final grades was employed. The results indicated that, under certain conditions, prediction of the final examination scores from the oral grades was possible; these conditions were as follows: first, the assessment mode producing the oral and final grades being written exams; and second, the other factors of the educational process (except the students of the course) remaining invariant. The credibility of the predictive models was verified. We also found that the assessment method used to produce the oral grades greatly affected the performance (validity) of the model. Hence, we estimated that it would not always be possible to produce predictive models. Our findings are in agreement with the results of other researchers [6,7,41].

The main findings of this study are summarized as follows:

- Under the assumption that the main factors of the teaching process such as the instructor, objectives, textbook, teaching style, assessment policy, etc., remain constant, it is possible to produce reliable predictive models;
- A model changes when one or more critical factor affecting the teaching process change;
- Different courses produce different models;
- When alternative assessment was used to produce the oral grades, it was impossible to produce statistically significant models;
- It is useful to keep historical data per course, in order to build a predictive model.

Our results also provide a deeper insight into alternative assessment and could assist educators in choosing methods to enrich their oral assessment policy [44].

This research was exploratory in nature and was specifically limited to the undergraduate telecommunications and electronics engineering cadets of the Academy and the specific courses, as offered by their instructors. Further research is needed to identify the conditions that would allow us to generalize our findings.

Author Contributions: Conceptualization, A.A.; methodology, A.A. and A.L.; software, A.L. and A.A.; validation, A.L. and A.A.; resources, A.A. and A.L.; data curation, A.A. and A.L.; writing—original draft preparation, A.A.; writing—review and editing, A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was approved by the Institutional Review Board of the Hellenic Air Force Academy (protocol number E/2023), on 11 October 2023.

Informed Consent Statement: This study does not reveal personal student data.

Data Availability Statement: Data not available due to institutional privacy protection policy.

Acknowledgments: The authors wish to thank I. Lekea, G. P. Kladis and M. Smyrnakis for providing useful suggestions and bibliography.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Andreatos, A.S. Designing educational scenarios to teach network security. In Proceedings of the 2017 IEEE Global Engineering Education Conference (EDUCON), Athens, Greece, 25–28 April 2017; pp. 1606–1610. [CrossRef]
2. Janisch, C.; Liu, X.; Akrofi, A. Implementing Alternative Assessment: Opportunities and Obstacles. *Educ. Forum* **2007**, *71*, 221–230. [CrossRef]
3. Andreatos, A. An Educational Scenario for Teaching Cyber Security Using low-cost Equipment and Open Source Software. In Proceedings of the ECCWS 2023, 22nd European Conference on Cyber Warfare and Security, Athens, Greece, 22–23 June 2023.
4. Ramanathan, P.; Fernandez, E. Can Early-Assignment Grades Predict Final Grades in IT Courses? In Proceedings of the 2017 ASEE Annual Conference & Exposition, Columbus, OH, USA, 24 June 2017. [CrossRef]
5. Martins, M.V.; Tolledo, D.; Machado, J.; Baptista, L.M.T.; Realinho, V. Early Prediction of student's Performance in Higher Education: A Case Study. In *Trends and Applications in Information Systems and Technologies*; WorldCIST 2021, Advances in Intelligent Systems and Computing; Rocha, Á., Adeli, H., Dzemyda, G., Moreira, F., Ramalho Correia, A.M., Eds.; Springer: Cham, Switzerland, 2021; Volume 1365._16. [CrossRef]
6. Cui, Y.; Chen, F.; Shiri, A. Scale up predictive models for early detection of at-risk students: A feasibility study. *Inf. Learn. Sci.* **2020**, *121*, 97–116. [CrossRef]
7. Davison, C.B.; Dustova, G. A Quantitative Assessment of Student Performance and Examination Format. *J. Instr. Pedagog.* **2017**, *18*. Available online: https://aabri.com/manuscripts/162516.pdf (accessed on 4 July 2022).
8. Zajacova, A.; Lynch, S.M.; Espenshade, T.J. Self-Efficacy, stress and academic success in college. *Res. High. Educ.* **2005**, *46*, 677–706. [CrossRef]
9. McKenzie, K.; Schweitzer, R. Who Succeeds at University? Factors predicting academic performance in first year Australian university students. *High. Educ. Res. Dev.* **2001**, *20*, 21–33. [CrossRef]
10. Siang, J.J.; Santoso, H.B. Learning Motivation and Study Engagement: Do they correlate with GPA? An evidence from Indonesian University. *J. Arts, Sci. Commer.* **2016**, *VII*, 111–118. [CrossRef]
11. Renk, K.; Smith, T. Predictors of academic-related stress in college students: An examination of coping, social support, parenting, and anxiety. *Naspa J.* **2007**, *44*, 405–431. [CrossRef]
12. Abdullah, N.A.; Shamsi, N.A.; Jenatabadi, H.S.; Ng, B.-K.; Mentri, K.A.C. Factors Affecting Undergraduates' Academic Performance during COVID-19: Fear, Stress and Teacher-Parents' Support. *Sustainability* **2022**, *14*, 7694. su14137694. [CrossRef]

13. Sadoughi, M.; Hejazi, S.Y. Teacher support and academic engagement among EFL learners: The role of positive academic emotions. *Stud. Educ. Eval.* **2021**, *70*, 101060. [CrossRef]
14. Lei, H.; Cui, Y.; Chiu, M.M. The relationship between teacher support and students' academic emotions: A meta-analysis. *Front. Psychol.* **2018**, *8*, 2288. [CrossRef]
15. Varga, M. The Effects of Teacher-Student Relationships on the Academic Engagement of Students. Master's Thesis, Goucher College, Baltimore, MD, USA, 2017. [CrossRef]
16. Rimm-Kaufman, S.; Sandilos, L. Improving Students' Relationships with Teachers to Provide Essential Supports for Learning. April 2010. Available online: https://www.apa.org/education-career/k12/relationships (accessed on 8 September 2022).
17. Spilt, J.L.; Koomen, H.M.Y.; Thijs, J.T. Teacher Wellbeing: The Importance of Teacher-Student Relationships. *Educ. Psychol. Rev.* **2011**, *23*, 457–477. [CrossRef]
18. Skinner, E.; Greene, T. Perceived control, coping, and engagement. In *21st Century Education: A Reference Handbook*; Good, T.L., Ed.; SAGE Publications, Inc.: Thousand Oaks, CA, USA, 2008; Volume 2, pp. I121–I130. [CrossRef]
19. Gehlbach, H.; Brinkworth, M.E.; Harris, A.D. Changes in teacher-student relationships. *Br. J. Educ. Psychol.* **2012**, *82 Pt 4*, 690–704. [CrossRef] [PubMed]
20. Boynton, M.; Boynton, C. Developing positive teacher-student relationships. In *The Educator's Guide to Preventing and Solving Discipline Problems*; ASCD: Alexandria, VA, USA, 2005.
21. Hargis, C.H. *Grades and Grading Practices. Obstacles to Improving Education and to Helping At-Risk Students*, 2nd ed.; Thomas: Springfield, IL, USA, 2003.
22. Haladyna, T.M. *A Complete Guide to Student Grading*; Allyn and Bacon: Needham Heights, MA, USA, 1999.
23. Tuckman, B.W. Using tests as an incentive to motivate procrastinators to study. *J. Exp. Educ.* **1998**, *66*, 141–147. [CrossRef]
24. Tuckman, B.W. The relative effectiveness of incentive motivation and prescribed learning strategies in improving college students' course performance. *J. Exp. Educ.* **1996**, *64*, 197–210. [CrossRef]
25. Schunk, D.H. Self-efficacy and Academic Motivation. *Educ. Psychol.* **1991**, *26*, 207–231. [CrossRef]
26. Covington, M.V. Goal theory, motivation, and school achievement: An integrative review. *Annu. Rev. Psychol.* **2000**, *51*, 171–200. [CrossRef] [PubMed]
27. Chen, M.H.; Liao, J.L. Correlations among Learning Motivation, Life Stress, Learning Satisfaction, and Self-Efficacy for Ph.D Students. *J. Int. Manag. Stud.* **2013**, *8*, 157–162.
28. Salanova, M.; Schaufeli, W.; Martínez, I.; Bresó, E. How obstacles and facilitators predict academic performance: The mediating role of study burnout and engagement. *Anxiety Stress Coping* **2010**, *23*, 53–70. [CrossRef]
29. Rasul, S.; Bukhsh, Q. A study of factors affecting students' performance in examination at university level. *Procedia-Soc. Behav. Sci.* **2011**, *15*, 2042–2047. [CrossRef]
30. Kumari, A.; Jain, J. Examination Stress and Anxiety: A Study of College Students. *Glob. J. Multidiscip. Stud.* **2014**, *4*, 31–40.
31. Linn, R.L. A century of standardized testing: Controversies and pendulum swings. *Educ. Assess.* **2001**, *7*, 29–38. [CrossRef]
32. Sawchuk, S. Is It Time to Kill Annual Testing? 8 January 2019. Available online: https://www.edweek.org/teaching-learning/is-it-time-to-kill-annual-testing/2019/01?intc=EW-BIG-NXT (accessed on 12 March 2023).
33. Bernal-Morales, B.; Rodríguez-Landa, J.F.; Pulido-Criollo, F. *Impact of Anxiety and Depression Symptoms on Scholar Performance in High School and University Students, a Fresh Look at Anxiety Disorders*; IntechOpen: London, UK, 2015.
34. Chapell, M.S.; Blanding, Z.B.; Silverstein, M.E.; Takahashi, M.; Newman, B.; Gubi, A.; McCann, N. Test anxiety and academic performance in undergraduate and graduate students. *J. Educ. Psychol.* **2005**, *972*, 268–274. [CrossRef]
35. Struthers, C.W.; Perry, R.P.; Menec, V.H. An examination of the relationships among academic stress, coping motivation, and performance in college. *Res. High. Educ.* **2000**, *41*, 581–592. [CrossRef]
36. Pascoe, M.C.; Hetrick, S.E.; Parker, A.G. The impact of stress on students in secondary school and higher education. *Int. J. Adolesc. Youth* **2019**, *25*, 104–112. [CrossRef]
37. Pope, D. 7 Approaches to Alternative Assessments. *ASCD 5 min.* 2019, 15. Available online: https://www.ascd.org/el/articles/7-approaches-to-alternative-assessments (accessed on 22 March 2023).
38. Shahiri, A.M.; Husain, W.; Rashid, N.A. A review on predicting student's performance using data mining techniques. *Procedia Comput. Sci.* **2015**, *72*, 414–422. [CrossRef]
39. El Aissaoui, O.; El Alami El Madani, Y.; Oughdir, L.; Dakkak, A.; El Allioui, Y. A Multiple Linear Regression-Based Approach to Predict Student Performance. In *Advanced Intelligent Systems for Sustainable Development (AI2SD'2019), Advances in Intelligent Systems and Computing*; Ezziyyani, M., Ed.; Springer: Cham, Switzerland, 2020; Volume 1102. [CrossRef]
40. Aldowah, H.; Al-Samarraie, H.; Fauzy, W.M. Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telemat. Inf.* **2019**, *37*, 13–49. [CrossRef]
41. Alabbad, J.; Abdul Raheem, F.; Almusaileem, A.; Almusaileem, S.; Alsaddah, S.; Almubarak, A. Medical students' logbook case loads do not predict final exam scores in surgery clerkship. *Adv. Med. Educ. Pract.* **2018**, *18*, 259–265. [CrossRef]
42. Chung J.-M., "Introduction to TCP/IP", MOOC Offered by Yonsei University, Korea. Available online: https://www.coursera.org/learn/tcpip (accessed on 12 June 2023).
43. Glen, S. Mean Squared Error: Definition and Example. Available online: https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/mean-squared-error/ (accessed on 23 August 2023).

44. Andreatos, A. Designing Alternative Assessment Activities and Adaptive Learning Scenarios to Cover Various Learning Styles in Higher Education. In *Fostering Pedagogy through Micro and Adaptive Learning in Higher Education: Trends, Tools, and Applications*; Queirós, R., Cruz, M., Pinto, C., Mascarenhas, D., Eds.; IGI Global: Hershey, PA, USA, 2023; pp. 287–305. [CrossRef]
45. Al-Saud, L.M. Learning style preferences of first-year dental students at King Saud University in Riyadh, Saudi Arabia: Influence of gender and GPA. *J. Dent. Educ.* **2013**, *77*, 1371–1378. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.