Blending E-Learning with Hands-on Laboratory Instruction in Engineering Education

An Experimental Study on Early Prediction of Student Performance and Behavior

https://doi.org/10.3991/ijet.v17i20.33141

Angelos Charitopoulos[1](✉), Maria Rangoussi[1], Dimitrios Koulouriotis[2]
[1]University of West Attica, Athens-Egaleo, Greece
[2]Democritus University of Thrace, Xanthi, Greece
acharito@uniwa.gr

Abstract—Among the various information sources exploited for the improvement of the learning process and outcomes, access and usage data from the interaction of students with e-learning platforms along with (past) student performance data are established as the two most meaningful and informative groups of variables. In the present study, these two groups of variables are jointly investigated as to their efficiency in providing both accurate and early prediction of student performance and behavior. The relevant educational intervention is designed and implemented as a quasi-experiment with undergraduate Electrical and Electronics Engineering students, under a novel approach that blends e-learning (asynchronous e-study and synchronous e-assessment) with a hands-on laboratory component. Can educational data mining algorithms provide both early and accurate prediction of student performance and student behavior under this scenario? If yes, how much prediction accuracy can be traded for prediction timeliness in order to allow a proactive class instructor take supportive measures for weak/marginal students, implementing a 'self-contained' strategy? To answer these questions, real data from the interaction of 3 academic year student cohorts with moodle are collected and analyzed. Results reveal that the proposed scenario can afford both accurate and early prediction of student performance and behavior, on the basis of data collected within the running academic term. The middle of the term is indicated as the earliest time point for getting meaningful predictions. Moreover, clustering of the data in the selected feature space reveals a consistent and therefore exploitable behavior of students along the term.

Keywords—e-learning, e-assessment, engineering education, hands-on laboratory, educational data mining, student performance, prediction, regression, clustering

1 Introduction

Long before the urgent Covid-19 pandemic condition, e-learning has emerged as a technically viable solution for the education of student cohorts of practically any size, thanks to the advances in Information and Communication Technologies (ICT). Intensive research has been carried out during the last three decades to clarify the role of this new paradigm and to quantify its impact on the learning outcomes as well as on the learners, the instructors and the whole community of education (see, e.g., [1]).

E-learning is not well-suited to hands-on laboratories, however. Engineering education relies heavily on hands-on laboratories to provide quality education and practical experience to young engineers. Hands-on laboratory modules or sessions have traditionally been included as mandatory components in undergraduate curricula across practically all engineering disciplines (e.g., [2]). Accreditation bodies for engineering education curricula include relevant items in their accreditation criteria – see, e.g. the American ABET [3] or the Canadian relevant body [4] – because hands-on labs are considered as beneficial for the students in multiple ways. As a result, e-learning has found limited use in the context of engineering education up to now, mostly as an auxiliary, asynchronous source of material for the students, to study and get prepared for the lab sessions.

The approach proposed and tested in this paper blends e-learning with a hands-on lab component, in the context of an Electrical and Electronics Engineering curriculum, by focusing the e-learning component to the (synchronous) student assessment task [5]. In addition, students access and use the e-learning platform asynchronously before each lab session, in order to study the course material and get prepared (e-study).

E-assessment has been proposed and put to use for the certification of academic or professional qualifications via online testing, either remote or on campus [6], [7] and, more recently, as the sole available means of student evaluation during the Covid-19 pandemic (e.g., [8]). Today, e-assessment is gradually gaining acceptance thanks to newer technologies that address its recognized issues such as reliability, security and technical soundness (EAA, n.d.). An advantage of e-assessment that is particularly welcome from a pedagogical aspect is that it affords instant and instructive/constructive feedback to the student, either at the class or at the individual level, [9], while it offers the instructor a variety of functionalities for the pedagogically correct handling of student errors.

The proposed approach constitutes a special case of blended or hybrid learning [10], [11]. Its major advantage is that the nature of the lab as a hands-on, practical source of experience for the student remains intact, while e-assessment adds to the objectivity of student evaluation. Traditionally, student evaluation is performed by the lab instructor on the basis of the practical aptitude exhibited by the students – something that is not easily quantifiable. Written tests are often employed to add objectivity to the lab evaluation. In the proposed scenario, these tests are taken online, in class, in the *moodle* e-learning platform. Although the hands-on laboratory that serves as the test bed for the present research is organized as a collaborative learning activity [12], where students work in small teams of 2 or 3 and produce a common report of their lab work each week, e-assessment is performed on a personal basis because an accurate estimate of the individual performance and the skills mastered is necessary for fair grading.

2 Educational Data Mining for the early prediction of student performance

Educational Data Mining (EDM) is applied to the data collected during e-learning/e-assessment in order to get answers to education related questions; the ultimate goal is to support informed decisions and measures that will eventually improve the experience and outcomes of education [13]–[17]. Major aims of EDM have been identified as the assessment and prediction of student performance, the monitoring of student dedication, the prediction of attrition (dropout rates) and the automated generation of recommendations for students or teachers (see, e.g., [18], [19]). Artificial Intelligence algorithms are increasingly being exploited to achieve these aims [20], [21].

In practice, however, the prediction of student performance is useful and therefore meaningful only if it is both accurate and early. Such a prediction would serve as an 'early warning system' that would allow the class instructor(s) enough time to take supportive measures for weak/marginal performance students and also allow for the measures to produce results before the completion of the study period. On the other hand, any prediction method applied on the output of a dynamic system, such as education, is plagued by the controversy that if one waits 'long enough' to get 'enough' data, an almost 100% accurate prediction is possible, but the window of opportunity for the exploitation of such prediction, unfortunately, will have expired by the time the prediction becomes available. This is especially true in higher education, where (i) courses are organized in short terms, typically semesters or trimesters, while (ii) any supportive measures need a reasonable time before they bear fruit. Despite the obvious usefulness of a really early warning system (see, e.g., research question 2 in [22], the majority of relevant research works focus on prediction accuracy instead of prediction timeliness – and they do so with excellent results (see, e.g., [13]–[16], and the numerous reviewed references therein). Regarding the time axis, dominant approaches in these studies are (i) prediction of the final outcome (e.g., graduation grade) of the whole study program – an event expected to come some 2 or 3 years after the time of prediction, (ii) prediction at the beginning of the 1st academic year of the student performance (or the dropout probability) to be attained by the end of the same year, (iii) prediction of course failure before even taking the course, [23]. These approaches afford 'long windows' of one or more academic years for the pedagogical utilization of the predicted values.

Among the limited number of studies that focus on the issue of timeliness, [24] uses a dynamic, longitudinal perspective and adopt a weekly basis for their analysis, although the authors' aim is not to develop an early warning system; [25] does develop such a system that uses three time points in the semester (weeks 4, 8 and 13) and compare prediction accuracies, for a purely online learning setting; [26] develops and, furthermore, puts to use such a system on week 4 of a 14 week long course, under a blended learning scenario. Their purely linear approach, however, leaves around 40% of the variability of the final outcome unexplained.

In contrast to these approaches, the present study focuses on short-term prediction that could be both obtained and utilized within the same course and the same academic term (e.g., an acedemic semester), under the proposed hands-on lab – e-study – e-assessment blended scenario. To this end, linear and non-linear algorithms are comparatively evaluated on the basis of longitudinal data of two types: (i) e-learning system access and usage data, (ii) current student performance data obtained in the same lab. Data should be put to use (fed to the corresponding algorithm) as soon as it becomes available, while the academic term is in progress. A self-contained strategy that does not rely on external data such as demographics or SAT scores or grades obtained in earlier years by the student through this or other e-learning systems, but collects and analyzes its own, pertinent data and utilizes results within its own time span of operation is both meaningful and practical – at least until e-learning platforms become mature enough to share, combine and exploit data across modules and courses.

The self-contained strategy is meaningful especially in the case of hands-on labs because practical skills are sought and evaluated therein. Prediction of student performance that is accurate as regards other components of the study program may therefore prove inaccurate for a hands-on lab: students predicted not in need of support by variables such as demographics or previous grades, may prove in fact to need support in the lab, while others predicted as weak may prove in fact to do well or even to excel in the lab.

The question that arises under this context is how much prediction accuracy can be traded in order to shift the prediction point early enough in the semester, and whether such a trade-off would still produce meaningful results for the instructor. More specifically, can the proposed scenario produce early and reliable predictions of the student progress and final outcome in the lab (a) at the beginning of the term, and (b) at the middle of the term?

3 Methods and tools

A quasi-experimental study has been planned and carried out in the form of educational interventions during three consecutive academic years, in an 4th-year undergraduate course on Digital Signal Processing (DSP), in a 5-year Electrical and Electronics Engineering curriculum. The DSP course includes a hands-on lab running once per week for a full academic semester (13 weeks) each year. In the lab, students use the Texas Instruments TMS320C5504/05™ Digital Signal Processor mounted on a Texas Instruments board to program and run a series of digital audio tasks, such as filters for specific audio effects reproduced digitally in real time. The moodle e-learning platform is used both for the asynchronous part (e-study) and the synchronous part (e-assessment).

The set of features or variables shown in Table 1, related to (i) the interaction of the students with the moodle platform (access and usage data) and (ii) the current performance of the students in the lab (grades obtained on weekly e-assessment tests), have been selected for collection, extraction and analysis because they are both intuitively meaningful/informative and practically accessible without the need to resort to external sources:

Table 1. Features or variables selected for collection and analysis

1	*Clicks*	the number of selections made by the user, using the mouse/pointing device
2	*Scrolls*	the number of computer screen scrolls done by the user while he/she navigates in the content
3	*Page Loads*	the number of page loads (requests to load a new page) done by the user did while he/she navigates in the content
4	*Nominal Time*	the total time a user is connected to the platform
5	*Active Time*	the time a user remains active while being connected to the platform, (typically shorter than nominal time, because the latter includes leisure/inactive time as well)
6	*Test Grade 1, …, Test Grade 8*	the grade obtained by the student in each of the 8 e-assessment tests taken during the semester
7	*Grade*	the final grade obtained by the student in the lab, computed as a weighted average of partial grades obtained in the various activities (a numerical value in the 0 to 10 scale)
8	*Success*	the final outcome for the student in the lab, obtained at the end of the term (binary: Fail/Pass)

In accordance with the self-contained strategy adopted here, demographic variables or variables related to the student records of past academic performance are deliberately ommitted.

The extraction of pertinent access and usage data from e-learning platforms is not an automated process. A custom plug-in for moodle has been coded and embedded in the platform, to extract the specific set of variables and to format and preprocess them for further analysis, as detailed in [27]. It has been put to pilot use with the student cohorts of academic years 2015–16 and 2016–17, for verification purposes; preliminary results can be found in [28], [29].

Subsequently, the data extraction process has run for three consecutive academic years (fall semesters) right before the Covid-19 pandemic condition, namely, 2017–18, 2018–19 and 2019–20, at the end of each academic semester. The observation set thus gathered consists of $N_0 = 219$ observations, i.e. unique students that enrolled in the DSP course, identified by their enrollment ID numbers.

The set of 15 variables defined earlier, *{Clicks, Scrolls, Page loads, Nominal time, Active time, 8 Test grades, Grade, Success}*, have been computed after appropriate pre-processing for each observation (student). Furthermore, as the supervised learning algorithms require disjoint training and test sets, the complete data set is split in two parts: Part I consisting of $N_1 = 169$ observations (first two academic years cohorts) is used for training and Part II consisting of the rest $N_2 = 50$ observations (third academic year cohort) is used for testing.

Analysis is carried out in the WEKA open source software, developed in Java by the Waikato University, New Zealand and available under the GNU General Public License.

Three Tasks have been set up to answer the research question, as in Table 2:

Table 2. Tasks set up and tools employed

Task	Task Description	WEKA Tools Employed
1	Prediction of the final outcome for any given student (binary, Fail/Pass): a two-class *classification problem*	(i) Logistic Regression, (ii) MLP Classifier, (iii) RBF Classifier, (iv) the J48-consolidated Decision Tree algorithm
2	Prediction of the final grade for any given student (numerical, 0 to 10 scale): a *regression problem*	(i) Linear Re-gression, (ii) MLP Network Regressor, (iii) RBF Network Regressor
3	Clustering of students according to their performance: a *clustering problem* with a predefined number of clusters	(i) Clustering by k-means algorithm with k=3 clusters and Euclidean distance

Two milestones have been set (a) at the beginning of the term and (b) at the middle of the term, as these points would allow for a reasonable time window during which the instructor(s) may exploit results and take measures or adopt strategies to reverse those student paths predicted to lead to failure or bad results. As at the exact beginning of the term no data is expected to be available on any of the selected features, the first time point ('beginning of the term') is set not exactly at the beginning of the term but at an early time during the term (after one week of classes).

An initial run of linear regression of *Grade*, considered as the dependent variable, on a limited set of independent variables of the platform access and usage type only, namely, *{Clicks, Scrolls, Page loads, Active time}*, has yielded a correlation coefficient of $r = 0.5961$. This is an indication of a moderate underlying linear relation that allows a significant part of the variability of the dependent variable (*Grade*) to be explained by non-linear relations to these independent variables. Artificial Neural Network algorithms, Decision Tree algorithms and Non-linear regression algorithms are promising candidates for the investigation of these relations. As a consequence, all algorithms employed for the 3 Tasks outlined above are of non-linear character, with the single exception of Linear Regression used in Task 2.

The Multilayer Perceptron (MLP) and the Radial Basis Function (RFB) ANNs have been singled out of the set of available ANNs, as two of the most popular and widely applied network types. In Task 1 these ANNs are used as 2-class classifiers, the (binary) output variable being *Success*, while in Task 2 they are used as regressors, the (numerical) output variable being *Grade*. Squared error minimization algorithms of the gradient descent, the BFGS algorithm in specific, [30], are used in the training phase in both networks types.

4 Data analysis and results

4.1 Task 1: Prediction of the binary (Fail/Pass) final outcome in the lab, for any given student

Results on Task 1 are summarized in Table 3 in the form of 2 × 2 confusion matrices and classification scores per case. Because the purpose of prediction is to identify the weak students that need support, class "a" corresponds to 'Fail' (positive, student needs support) while class "b" corresponds to 'Pass' (negative, student does not need support).

Apart from the standard measures of Accuracy, Precision, Sensitivity and Specificity, the F1-score is also computed, given that (i) the two classes "a" and "b" are not equivalent, class "b" being considerably more numerous than "a", while (ii) type-II errors (false negatives, i.e. students predicted not to need help when in fact they do need it) are considered here more grave than type-I errors (false positives, i.e., students predicted to need help when in fact they don't need it).

Table 3. Results of Task 1: prediction of the binary (Fail/Pass) final outcome in the lab (*)

Item1	Beginning of the Term	Middle of the Term	End of the Term
Logistic Regression	a b <- classified as 8 2 | a = Fail 2 38 | b = Pass	a b <- classified as 9 1 | a = Fail 2 38 | b = Pass	a b <- classified as 10 0 | a = Fail 0 40 | b = Pass
	AC = 46/50 = 92.00% PR = 8/10 = 80.00% SE = 8/10 = 80.00% SP = 38/40 = 95.00% F1-score = 80.00%	AC = 47/50 = 94.00% PR = 9/11 = 81.81% SE = 9/10 = 90.00% SP = 38/40 = 95.00% F1-score = 85.70%	AC = 50/50 = 100.00% PR = 10/10 = 100.00% SE = 10/10 = 100.00% SP = 40/40 = 100.00% F1-score = 100.00%
MLP Network (Classifier)	a b <-- classified as 8 2 | a = Fail 1 39 | b = Pass	a b <-- classified as 8 2 | a = Fail 1 39 | b = Pass	a b <-- classified as 9 1 | a = Fail 0 40 | b = Pass
	AC = 47/50 = 94.00% PR = 8/9 = 88.88% SE = 8/10 = 80.00% SP = 39/40 = 97.50% F1-score = 84.20%	AC = 47/50 = 94.00% PR = 8/9 = 88.88% SE = 8/10 = 80.00% SP = 39/40= 97.50% F1-score = 84.20%	AC = 49/50 = 98.00% PR = 9/9 = 100.00% SE = 9/10 = 90.00% SP = 40/40 = 100.00% F1-score = 94.73%
RBF Network (Classifier)	a b <- classified as 7 3 | a = Fail 1 39 | b = Pass	a b <- classified as 9 1 | a = Fail 2 38 | b = Pass	a b <- classified as 9 1 | a = Fail 0 40 | b = Pass
	AC = 46/50 = 92.00% PR = 7/8 = 87.50% SE = 7/10 = 70.00% SP = 39/40 = 97.50% F1-score = 77.77%	AC = 47/50 = 94.0% PR = 9/11 = 81.81% SE = 9/10 = 90.00% SP = 38/40 = 95.00% F1-score = 85.70%	AC = 49/50 = 98.00% PR = 9/9 = 100.00% SE = 9/10 = 90.00% SP = 40/40 = 100.00% F1-score = 94.73%
J48 (Consolidated) Decision Tree	a b <- classified as 9 1 | a = Fail 4 36 | b = Pass	a b <- classified as 9 1 | a = Fail 2 38 | b = Pass	a b <- classified as 9 1 | a = Fail 1 39 | b = Pass
	AC = 45/50 = 90.00% PR = 9/13 = 69.23% SE = 9/10 = 90.00% SP = 36/40 = 90.00% F1-score = 78.26%	AC = 47/50 = 94.00% PR = 9/11 = 81.81% SE = 9/10 =90.00% SP = 38/40 = 95.00% F1-score = 85.70%	AC = 48/50 = 96.00% PR = 9/10 = 90.00% SE = 9/10 = 90.00% SP = 39/40 = 97.50% F1-score = 90.00%

Note: *Abbreviations explanation: AC = Accuracy, PR = Precision, SE = Sensitivity, SP = Specificity.

The four algorithms are presented in Table 3 in descending order of Accuracy in the *End of the term* column. If all three tabulated time points are considered, Logistic Regression yields the best scores in terms of Accuracy, followed by MLP and RBF neural networks that exhibit equivalent performances; J48 holds the last position.

It may be claimed that all examined classification algorithms produce satisfactory results (Accuracy, i.e., correct classification scores over 90%). If any single algorithm (any one row in Table 3) is examined, practically all measures increase with time and reach their peak values at the end of the term. It is interesting that all four methods yield comparable measures at the middle of the term.

4.2 Task 2: Prediction of the (numerical) final grade in the lab, for any given student

Results for linear and non-linear regression methods employed in Task 2 are summarized in Table 4, in terms of (i) the set of independent variables *Grade* regresses on and (ii) the correlation coefficients and errors obtained.

Table 4. Results of Task 2: linear and non-linear regression of Grade on sets of independent variables, correlation coefficients and errors (*)

Item1	Beginning of the Term (4 Independent Variables)	Middle of the Term (8 Independent Variables)	End of the Term (12 Independent Variables)
	Grade = *f* { Clicks, Scrolls, Page loads, Active Time}	*Grade* = *f* { Clicks, Scrolls, Page loads, Active Time, Test Grade-1, Test Grade-2, Test Grade-3, Test Grade-4}	*Grade* = *f* { Clicks, Scrolls, Page loads, Active Time, Test Grade-1, Test Grade-2, Test Grade-3, Test Grade-4, Test Grade-5, Test Grade-6, Test Grade-7, Test Grade-8}
Linear Regression	r = 0.5961 MAE = 1.15 RMSE = 1.52 RAE = 74.91% RRSE = 79.59%	r = 0.9337 MAE = 0.67 RMSE = 0.81 RAE = 44.09 % RRSE = 42.50%	r = 0.9936 MAE = 0.20 RMSE = 0.27 RAE =13.31% RRSE = 14.11%
MLP Network (Regressor)	r = 0.7883 MAE = 0.63 RMSE = 0.76 RAE = 41.30% RRSE = 39.97%	r = 0.9451 MAE = 0.63 RMSE = 0.76 RAE = 41.30% RRSE = 39.973%	r = 0.9922 MAE = 0.63 RMSE = 0.76 RAE = 41.30% RRSE = 39.97%
RBF Network (Regressor)	r = 0.7643 MAE = 0.92 RMSE =1.30 RAE = 60.28% RRSE = 67.97%	r = 0.9358 MAE = 0.72 RMSE = 0.87 RAE = 46.87% RRSE = 45.71%	r = 0.9924 MAE = 0.21 RMSE = 0.29 RAE = 14.10% RRSE = 15.32%

Note: *Abbreviations explanation: r = Correlation Coefficient, MAE = Mean Absolute Error, RMSE = Root Mean Square Error, RAE = Relative Absolute Error, RRSE = Root Relative Squared Error.

As it can be seen in Table 4, correlation coefficient r increases monotonically with time across all three methods. At the middle of the term, it has already reached the level of 0.93–0.94 approximately, method-dependent, which is very satisfactory.

4.3 Task 3: Clustering of the students at the middle and at the end of the term

The number of clusters k is set to k = 3, by an experimental search in the range of k = 2 to k = 10. Three alternative methods for the selection of the optimal k are used, namely (i) Calinski-Harabasz [31], (ii) Silhouette [32], and (iii) Davies-Bouldin [33]. The first two are maximization methods (the criterion is maximized on the optimal k value) while the third one is a minimization method (the criterion is minimized on the optimal k value). Results shown in Figure 1 indicate that k = 3 is the optimal cluster number decided unanimously by the three criteria.

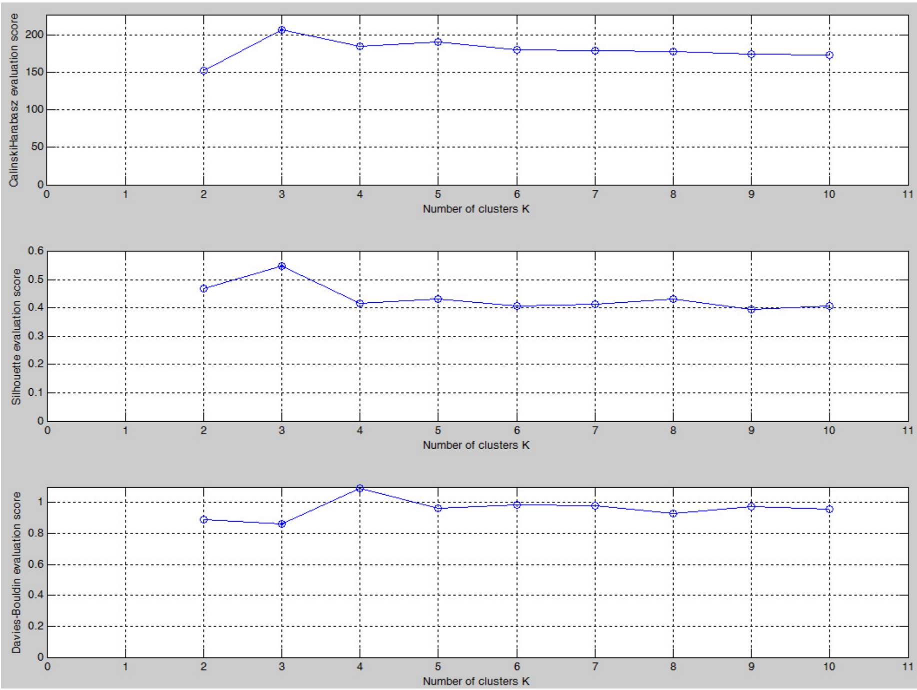

Fig. 1. Three alternative optimization criteria for the selection of the number of clusters k, evaluated in the range of k = 2 to k = 10: Calinski-Harabasz (top curve – maximization), Silhouette (middle curve – maximization), Davies-Bouldin (bottom curve – minimization)

At the middle of the term, data is available on 9 out of the 15 features, as these are defined in Section 2, namely, on *{Clicks, Scrolls, Page Loads, Nominal Time, Active Time, Test Grade-1, Test Grade-2, Test Grade-3, Test Grade-4}*, while at the end of the term data on 6 more features becomes available, namely, on *{Test Grade-5, Test Grade-6, Test Grade-7, Test Grade-8, Success, Grade}*. Clustering may be performed either on the multi-dimensional space of all these features or on any meaningful subset of them. In terms of the visualization and interpretation of the results, however, clustering on the basis of only two variables (features) is advantageous, because the results can be shown in a 2D scatter plot which allows for direct visual evaluation and are more easily comprehensible and interpretable. The two features selected here are (i) the average grade obtained at the middle of the semester (average over the first 4 test grades) and (ii) the active time dedicated by the student to the course up to the same time point.

These two features are considered both meaningful and informative, as they connect student dedication to performance.

The k-means algorithm is used for clustering and the Euclidean distance is selected as the vector distance measure. Clustering results are shown in Figure 2, in the form of centroids of each cluster and the population assigned to it, and in Figure 3, in the form of a 2-D scatter plot on the (*average Grade at the middle of the term, Active time*) axes.

```
Final cluster centroids:
                                    Cluster#
Attribute          Full Data            1          3          2
                     (219.0)       (43.0)     (63.0)    (113.0)
===========================================================
MID_ACTIVE_TIME     166.2294      349.096    96.5366   135.4982
MID_TERM_GRADE        5.1236       5.9393     3.1506     5.9131

Time taken to build model (full training data) : 0.03 seconds

=== Model and evaluation on training set ===
Clustered Instances
1        43 ( 20%)
3        63 ( 29%)
2       113 ( 52%)
```

Fig. 2. Clustering results at the middle of the term (N = 219 students, k = 3 clusters, k-means, Euclidean distance)

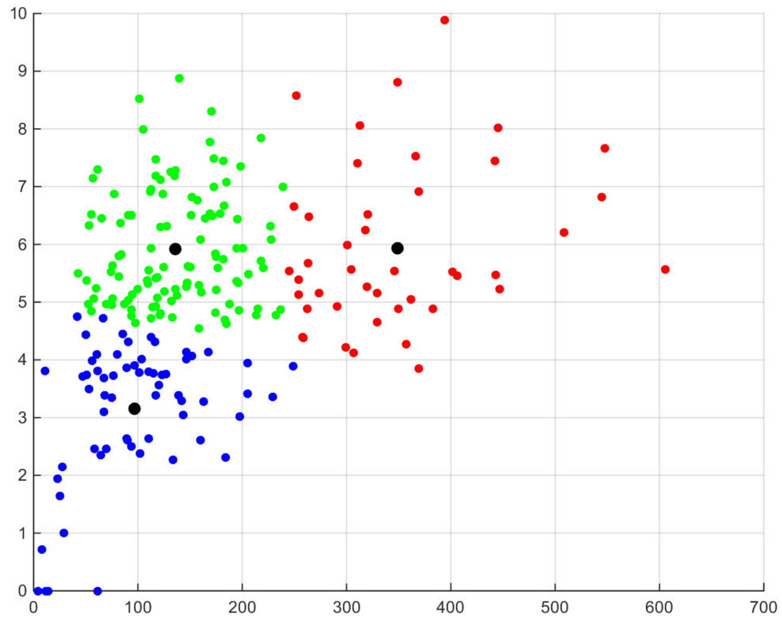

Fig. 3. Clustering scatter plot at the middle of the term (N_0 = 219 students, k = 3 clusters, k-means, Euclidean distance). Vertical axis: Average grade at midterm (0 to 10 scale); horizontal axis: active time dedicated to the course up to midterm (minutes). Cluster 1 (red), Cluster 2 (green), Cluster 3 (blue). Cluster centroids are depicted as black dots

Three different, well-defined student behaviors emerge clearly from the visual evaluation of the clustering results in Figure 3:

(i) Cluster 1 (in red, upper right quadrant in Figure 3), represented by Centroid (349.096, 5.9393): the 43 students (20%) clustered here combine high dedication with high grades,
(ii) Cluster 2 (in green, upper left quadrant in Figure 3), represented by Centroid (135.4982, 5.9131): the 113 students (52%) clustered here combine medium dedication with high grades,
(iii) Cluster 3 (in blue, lower left quadrant in Figure 3), represented by Centroid (96.5366, 3.1506): the 63 students (29%) clustered here combine low-to-medium dedication with low grades.

An interesting outcome is that the lower right quadrant in Figure 3 (high dedication and low grades) is almost empty. Indeed, it would be counter-intuitive if high dedication would result in low performance.

The effectiveness of clustering attempted at the middle of the term is validated by repeating the same type of clustering at the end of the term, when data is available on all 15 features. The two features used are *Grade* (now it is computed as the average of all 8 *Test grades*) and the total *Active time* recorded for each student at the end of the term. Results are shown in Figure 4, in the form of centroids of each cluster and the population assigned to it, and in Figure 5, in the form of a 2-D scatter plot on the (*Grade*, *Active time*) axes.

```
Final cluster centroids:
                              Cluster#
Attribute          Full Data         2          3          1
                     (219.0)    (105.0)     (75.0)     (39.0)
============================================================
Total_Active_Time   277.0489  254.1780   142.1177   598.1077
Grade                 4.9243    5.7391     3.1703     6.1036

Time taken to build model (full training data) : 0.02 seconds

=== Model and evaluation on training set ===
Clustered Instances
2      105 ( 48%)
3       75 ( 34%)
1       39 ( 18%)
```

Fig. 4. Clustering results at the end of the term (N_0 = 219 students, k = 3 clusters, k-means, Euclidean distance)

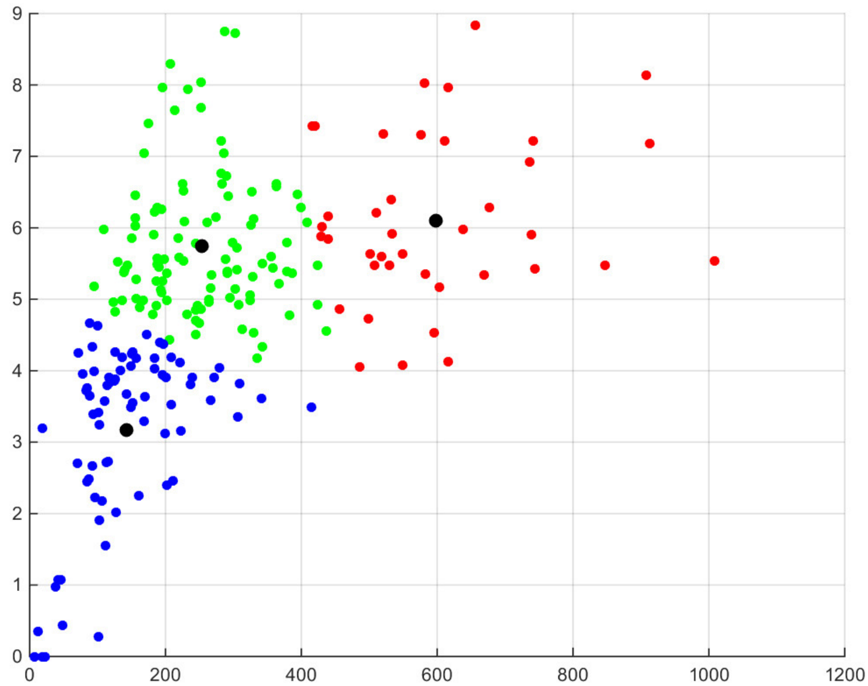

Fig. 5. Clustering scatter plot at the end of the term (N_0 = 219 students, k = 3 clusters, k-means, Euclidean distance). Vertical axis: Average grade at the end of the term (0 to 10 scale); horizontal axis: active time dedicated to the course up to the end of the term (minutes). Cluster 1 (red), Cluster 2 (green), Cluster 3 (blue). Cluster centroids are depicted as black dots

The three different student behaviors suggested by Figure 3 (middle of the term) are essentially repeated in Figure 5 (end of the term):

(i) Cluster 1 (in red, upper right quadrant in Figure 5), represented by Centroid (598.1077, 6.1036): the 39 students (18%) clustered here combine high dedication with high grades,
(ii) Cluster 2 (in green, upper left quadrant in Figure 5), represented by Centroid (254.1780, 5.7391): the 105 students (48%) clustered here combine medium dedication with high grades, and
(iii) Cluster 3 (in blue, lower left quadrant in Figure 5), represented by Centroid (142.1177, 3.1703): the 75 students (34%) clustered here combine low-to-medium dedication with low grades.

A detailed comparison of student cluster assignments at the middle and at the end of the term exposes a few cases of student transitions from cluster to cluster. Table 5 shows these transition cases from the cluster assigned at the middle of the term (vertical dimension in Table 5) to the cluster assigned at the end of the term (horizontal dimension in Table 5). Limited as they may be, the transitions are not counter-intuitive as they take place between adjacent clusters: cluster 1 (‘good students’) is slightly decreased during the semester (–9.0%), cluster 2 ('average students') exhibits a considerable mobility and is eventually also slightly decreased (–7.0%) while cluster 3 ('weak students') both gives students to and takes students from its adjacent cluster 2 and is eventually increased (+19.0%). In total, transitions amount to 38/ 219 cases or 17.35% – a percentage low enough to justify the use of the midterm results as a satisfactory estimate of the "true" clustering results computable only at the end of the term.

Table 5. Student transition matrix from the cluster assigned at the middle of the term (vertically) to the cluster assigned at the end of the term (horizontally). Total number of students $N_0 = 219$

	Item1	**Cluster Assigned at the End of the Term**			
		Cluster 1 (RED)	**Cluster 2 (GREEN)**	**Cluster 3 (BLUE)**	**Totals**
Cluster assigned at the middle of the term	Cluster 1 (RED)		4	0	4
	Cluster 2 (GREEN)	0		23	23
	Cluster 3 (BLUE)	0	11		11
	Totals	0	15	23	38

5 Interpretation of the results and discussion

As it has been experimentally verified in Tasks 1 and 2, prediction of the final outcome per student (*Success*) as well as prediction of the performance in the lab (*Grade*) can be obtained with very good accuracy at the middle of the term. Indeed,

- regarding *Success* (Fail/Pass prediction), all four methods compared yield Accuracy around 94% (see Table 3, middle column) – in fact, all the measures computed take on comparable values at the middle of the term,
- regarding *Grade*, both linear and non-linear regression methods yield correlation coefficients of around 94%; among them, the MLP Regressor produces the higher correlation coefficient (94.51%) (Table 4, middle column).

The corresponding results at the beginning of the term, however, need a more careful evaluation. *Success* predictions are at a very good level of accuracy, slightly lower than that at the middle of the term and yet high enough to allow the instructor a rough, 'binary' classification of the students already at the beginning of the term. Indeed, as it can be seen in Table 3 (left column), Accuracy ranges between 90% and 94%, method-dependent. This result is important because it indicates that, in practice, e-learning platform access and usage data alone can adequately support both accurate and early classification of the students, yet, only of this very rough, binary type.

On the contrary, the prediction of *Grade* attempted so early in the semester is not accurate. As it can be seen in Table 4 (*Beginning of the term* column), correlation coefficients range from 59.61% to 78.83%, method-dependent; again, the MLP regressor method produces the highest value (78.83%). As any experienced instructor would argue, this is intuitively meaningful because the final grade of any individual student is a value 'constructed' during the whole semester; therefore, it cannot be expected to be accurately predicted at the beginning of the term. This is in agreement with existing research establishing that (past) performance data is a necessary ingredient for good prediction of future performance, or as [24] put it: *'…the best predictor for performance, is performance itself*'.

The above results on *Success* and *Grade* are also in agreement with earlier research indicating that access and usage data collected from the interaction of students with learning management systems (i) may conditionally be useful variables as predictors of student achievement, (e.g., [34]), while at the same time (ii) are inadequate/ poor sources of information when more elaborate research questions are posed [24].

The results obtained in Task 3, on the other hand, reveal certain behaviors of the students that may be exploited by the class instructor in the second half of the term. Regarding dedication in relation to performance, the same three 'core' student behaviors emerge consistently via clustering at the middle and at the end of the term. Moreover, the three clusters-behaviors gather roughly the same percentage of the class population at both these time points, specifically,

- 20% of the students fall into cluster 1 (red in Figures 3 and 5) of the high dedication and high performance combination,
- 50% of the students fall into cluster 2 (green in Figures 3 and 5) of the medium dedication and high performance combination, while
- 30% of the students fall into cluster 3 (blue in Figures 3 and 5) of the low-to-medium dedication and low performance combination.

These results can aid the class instructor to take corrective measures as to the structure of the lab teams and the 'mixture of students' put together in each team, in the middle of the term. The specific measures depend on the instructor's educational and pedagogical scenario and aims.

From another aspect, the results of all three Tasks are of practical value for the class instructor, when he/she is revisiting the e-assessment tests for restructuring or in general improving them. Given the fact that *Success* as well as *Grade* predictions depend critically on the e-assessment tests put together and delivered by the instructor, it is of great practical interest for him/her to monitor the accuracy, sensitivity and F1-score of the prediction method employed, which is 'fed' with the results from the specific e-assessment tests, in order to restructure or reform the e-assessment material for a more accurate and sensitive prediction of the student categories. Along the same line, cluster centroids obtained in Task 3 may serve as indicators to aid the instructor adjust the level of difficulty and/or the time allowed for the various assignments of e-assessment and/or normalize the total class success rates along academic years.

6 Conclusions – further research

A blended learning scenario that combines e-study and e-assessment with the physical participation in a hands-on laboratory within an undergraduate engineering curriculum is proposed in this paper. It is applied and tested through educational interventions for three consecutive academic years in the hands-on laboratory of the Digital Signal Processing course module, in the 7th semester of a 5-year undergraduate Electrical and Electronics Engineering curriculum. The aim is to predict student performance and student behavior early enough in the academic term so as to allow the class instructor take supportive measures that will bear fruit in the same term.

Regarding the trade-off between accuracy and timeliness of the results, both prediction and clustering results obtained via EDM algorithms for the analysis of the collected data indicate that although the beginning of the semester is rather early, the middle of the semester is a suitable time point that combines very good performance prediction results (accuracy around 94% and sensitivity around 90%, as compared to the 100% achieved at the end of the semester) to very good clustering results that produce meaningful interpretations (e.g., 82.65% of the students retain their cluster/behavior up to the end of the term). Despite the peculiarity of the educational scenario examined here, the results obtained tend to agree with existing results obtained by research under scenarios that either employ pure e-learning or blend e-learning and in-class lecturing, as to the usefulness of user-platform and user-content interaction data and of (past) performance data. These findings imply that data mining algorithms applied on the access, usage and performance data dynamically collected during the e-study and e-assessment phases of the proposed scenario can yield reliable descriptive (clustering) and predictive (performance) results early enough to give the instructor the time to react and take supportive measures for weak students.

Furthermore, the results from data mining within the proposed scenario may be exploited in multiple ways by the class instructor. Along this line, our further research is directed towards the development of a recommendation system to aid the class instructor dynamically modify instruction parameters guided by the EDM results.

7 Acknowledgment

This research was supported by the Special Account for Research Grants of the University of West Attica.

8 References

[1] R. C. Clark and R. E. Mayer, *E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning*. San Francisco, CA: Pfeiffer & Company, 2011. https://doi.org/10.1002/9781118255971

[2] L. D. Feisel and A. J. Rosa, "The role of the laboratory in undergraduate engineering education," *Journal of Engineering Education*, vol. 94, no. 1, pp. 121–130, 2005. https://doi.org/10.1002/j.2168-9830.2005.tb00833.x

[3] ABET, "Criteria for accrediting engineering programs 2021–22," 2006. [Online]. Available: https://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-engineering-programs-2021-2022/. [Accessed 31 December 2021].

[4] Engineers Canada (n.d.), "Canadian engineering accreditation board questionnaire for evaluation of an engineering program 2021". [Online]. Available: https://engineerscanada.ca/accreditation/accreditation-resources/2022-2023-accreditation-cycle. [Accessed 31 December 2021].

[5] E-Assessment Association (n.d.). [Online]. Available: https://e-assessment.com/. [Accessed 31 December 2021].
[6] N. Alruwais, G. Wills, and M. Wald, "Advantages and challenges of using e-assessment," *International Journal of Information and Education Technology*, vol. 8, no. 1, pp. 34–37, 2018. https://doi.org/10.18178/ijiet.2018.8.1.1008
[7] T. Daradoumis, J. M. M. Puig, M. Arguedas, and L. Calvet Liñan, "Analyzing students' perceptions to improve the design of an automated assessment tool in online distributed programming," *Computers & Education*, vol. 128, pp. 159–170, 2019. https://doi.org/10.1016/j.compedu.2018.09.021
[8] N. Doğan, N. Kibrislioğlu Uysal, H. Kelecioğlu, and R. K. Hambleton, "An overview of e-assessment," *Hacettepe University Journal of Education*, vol. 35 (Special Issue), pp. 1–5, 2020. https://doi.org/10.16986/HUJE.2020063669
[9] A. P. Cavalcanti, A. Barbosa, R. Carvalho, F. Freitas, Y.-S. Tsai, D. Gašević, and R. F. Mello, "Automatic feed-back in online learning environments: A systematic literature review," *Computers & Education: Artificial Intelligence*, vol. 2, pp. 100027, 2021. https://doi.org/10.1016/j.caeai.2021.100027
[10] D. R. Garrison and N. D. Vaughan, *Blended Learning in Higher Education: Framework, Principles, and Guidelines*. San Francisco, CA: Jossey-Bass Higher and Adult Education Series, John Wiley & Sons, 2011.
[11] C. Dziuban, C. R. Graham, P. D. Moskal, A. Norberg, and N. Sicilia, "Blended learning: The new normal and emerging technologies," *International Journal of Educational Technology in Higher Education*, vol. 15, no. 3, 2018. https://doi.org/10.1186/s41239-017-0087-5
[12] P. Dillenbourg, "Collaborative learning: Cognitive and computational approaches", in *Advances in Learning and Instruction Series*, Bingley, UK: Emerald Book Publishing, 1999.
[13] G. Siemens and R. S. Baker, "Learning analytics and educational data mining: Towards communication and collaboration," In Proc. 2nd International Conference on Learning Analytics and Knowledge (LAK'12), Vancouver, BC, Canada, 2012, pp. 252–254. https://doi.org/10.1145/2330601.2330661
[14] C. Romero and S. Ventura, "Data mining in education," *WIREs Data Mining and Knowledge Discovery*, vol. 3, pp. 12–27, 2013. https://doi.org/10.1002/widm.1075
[15] A. Peña-Ayala, "Educational data mining: A survey and a data mining-based analysis of recent works," *Expert Systems with Applications*, vol. 41 (4 part 1), pp. 1432–1462, 2014. https://doi.org/10.1016/j.eswa.2013.08.042
[16] H. Aldowah, H. Al-Samarraie, and W. M. Fauzy, "Educational data mining and learning analytics for 21st century higher education: A review and synthesis," *Telematics and Informatics*, vol. 37, pp. 13–49, 2019. https://doi.org/10.1016/j.tele.2019.01.007
[17] M. H. Bin Roslan and C. J. Chen, "Educational data mining for student performance prediction: A systematic literature review (2015–2021)," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 17, no. 5, pp. 147–179, 2022. https://doi.org/10.3991/ijet.v17i05.27685
[18] A. Charitopoulos, M. Rangoussi, and D. Koulouriotis, "On the use of soft computing methods in educational data mining and learning analytics research: A review of years 2010–2018," *International Journal of Artificial Intelligence in Education*, vol. 30, no. 3, pp. 371–430, 2020. https://doi.org/10.1007/s40593-020-00200-8
[19] S. Rakic, N. Tasic, U. Marjanovic, S. Softic, E. Lüftenegger, and I. Turcin, "Student performance on an e-learning platform: Mixed method approach," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 2, pp. 187–203, 2020. https://doi.org/10.3991/ijet.v15i02.11646

[20] I. Roll and R. L. Wylie, "Evolution and revolution in artificial intelligence in education", *International Journal of Artificial Intelligence in Education*, vol. 26, no. 2, pp. 582–599, 2016. https://doi.org/10.1007/s40593-016-0110-3

[21] M. Imran, S. Latif, D. Mehmood, and M. S. Shah, "Student academic performance prediction using supervised learning techniques," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 14, no. 14, pp. 92–104, 2019. https://doi.org/10.3991/ijet.v14i14.10310

[22] L. Cagliero, L. Canale, L. Farinetti, E. Baralis, and E. Venuto, "Predicting student academic performance by means of associative classification," *MDPI Applied Sciences*, vol. 11, pp. 1420, 2021. https://doi.org/10.3390/app11041420

[23] A. Polyzou and G. Karypis, "Feature extraction for next-term prediction of poor student performance," *IEEE Trans. on Learning Technologies*, vol. 12, pp. 237–248, 2019. https://doi.org/10.1109/TLT.2019.2913358

[24] D. T. Tempelaar, B. Rienties, and B. Giesbers, "In search for the most informative data for feedback generation: Learning analytics in a data-rich context," *Computers in Human Behavior*, vol. 47, pp. 157–167, 2015. https://doi.org/10.1016/j.chb.2014.05.038

[25] Y. H. Hu, C. L. Lo, and S. P. Shih, "Developing early warning systems to predict students' online learning performance," *Computers in Human Behavior*, vol. 36, pp. 469–478, 2014. https://doi.org/10.1016/j.chb.2014.04.002

[26] A. Jokhan, B. Sharma, and S. Singh, "Early warning system as a predictor for student performance in higher education blended courses," *Studies in Higher Education*, vol. 44, no. 11, pp. 1900–1911, 2019. https://doi.org/10.1080/03075079.2018.1466872

[27] G. Sykamiotis, A. Charitopoulos, M. Rangoussi and D. Koulouriotis, "Extraction and presentation of access and usage data from an e-learning platform (moodle): Design and development of a software application," In Proc. 8th IEEE Global Engineering Education Conference (EDUCON 2017), Athens, Greece, 2017. https://doi.org/10.1109/EDUCON.2017.7942973

[28] A. Charitopoulos, M. Rangoussi, and D. Koulouriotis, "E-learning platform access and usage statistics through data mining: An experimental study in Moodle," In Proc. 9th Intl. Conf. of Education, Research and Innovation (ICERI'16), Seville, Spain, 2016. https://doi.org/10.21125/iceri.2016.1638

[29] A. Charitopoulos, M. Rangoussi, and D. Koulouriotis, "Educational data mining and data analysis for optimal learning content management," In Proc. 8th IEEE Global Engineering Education Conference (EDUCON 2017), Athens, Greece, 2017. https://doi.org/10.1109/EDUCON.2017.7942969

[30] J. Nocedal, "Updating quasi-Newton matrices with limited storage," *Mathematics of Computation*, vol. 35, pp. 773–782, 1980. https://doi.org/10.1090/S0025-5718-1980-0572855-7

[31] T. Calinski and J. Harabasz, "A dendrite method for cluster analysis," *Communications in Statistics*, vol. 3, no. 1, pp. 1–27, 1974. https://doi.org/10.1080/03610927408827101

[32] P. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, no. 1, pp. 53–65, 1987. https://doi.org/10.1016/0377-0427(87)90125-7

[33] D. Davies and D. Bouldin, "A cluster separation measure," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 1, no. 2, pp. 224–227, 1979. https://doi.org/10.1109/TPAMI.1979.4766909

[34] L. P. Macfadyen and S. Dawson, "Mining LMS data to develop an 'early warning system' for educators: A proof of concept," *Computers and Education*, vol. 54, no. 2, pp. 588–599, 2010. https://doi.org/10.1016/j.compedu.2009.09.008

9 Authors

Angelos Charitopoulos (BSc, MSc) is a PhD candidate in the Department of Production and Management Engineering, Democritus University of Thrace, Xanthi, Greece and a member of the staff in the Department of Electrical and Electronics Engineering, University of West Attica, 250, Thivon str., GR-12241, Athens-Egaleo, Greece (e-mail: acharito@uniwa.gr). He is an active member of the Research Lab on Electronics and Computer Technologies in the same Department. His research interests are in Networking and Data Communications, Soft Computing Algorithms, Educational Technologies and E-learning.

Maria Rangoussi (MEng, MSc, PhD) is Professor in the Department of Electrical and Electronics Engineering, University of West Attica, 250, Thivon str., Athens-Egaleo, GR-12241, Greece (e-mail: mariar@uniwa.gr). She is the director of the MSc by Research in Electrical and Electronics Engineering in the same Department. Her research interests are in Digital Signal Processing and Pattern Recognition algorithms and their applications in various fields, as well as in Educational Technologies, Educational Data Mining and Learning Analytics. Prof. Rangoussi is a member of IEEE, ISCA and the Technical Chamber of Greece.

Dimitrios E. Koulouriotis (MEng, MSc, PhD) is Professor in Intelligent Systems in Industrial Engineering and Management and Head of the Department of Production and Management Engineering, in the Democritus University of Thrace, 12, Vas. Sofias str., GR-67100, Xanthi, Greece (e-mail: jimk@pme.duth.gr). He is also the Director of the Laboratories of a) Industrial Production and b) Ergonomics and Safety. His teaching and research activities are in Intelligent Systems, Computational Intelligence and Machine Learning, Industrial Engineering, Engineering and Project Management, Safety and Reliability, Innovation and Technology Management.

Article submitted 2022-06-09. Resubmitted 2022-07-20. Final acceptance 2022-07-21. Final version published as submitted by the authors.