To Predict Students' Success in the Next Class, We Need to Go Beyond (Reliable) Grades from the Previous Class: An Empirical Study

Annette M. Siemssen and Olga Kosheleva
Department of Teacher Education
University of Texas at El Paso
El Paso, Texas 79968, USA
amsiemssen@miners.utep.edu, olgak@utep.edu

Abstract

In a two-class sequence, it is important to be able to make sure that students graduating from the first class can succeed in the second one. If the cut-off for success in the first class is set too low, many ill-prepared students are allowed to take the second class and are thus doomed to fail it. If this cut-off is set too high, medium-prepared students who could potentially succeed in the next class waste time by unnecessarily repeating the first class. From this viewpoint, it is desirable to be able to predict the student's success in the second class based on this student's (reliable) grades in the first class. On the example of a two-class introductory computer science sequence, we show that in some situations, a reliable prediction is not possible. Namely, to get a good prediction, in addition to (reliable) grades for the exams, grades that reflect the students' ability to solve simple problems, we also need to take into account less reliable (and more cheating-prone) grades on take-home assignments such as labs, grades that reflect the students' ability to solve complex problems.

1 Introduction

Need to predict students' success. For many classes, the main objective is to prepare a student for the next class. Usually, if a student has a passing grade in the previous class, this student is eligible to sign up for the next class. Otherwise, if the student's grade in the first class is unsatisfactory, the student has to repeat this class – or, if the student has already tried that several times and failed, the student is dropped from the program.

In many cases, this procedure works well. Usually, when a student shows excellent or very good knowledge of the material from the first class, this clearly indicates that the student is ready for the following class. On the other hand, if the student's results in the first class are bad, this student is clearly not ready for the second class.

However, in borderline cases, it is often not easy to predict the student's success: sometimes, a student with a barely passing grade in the first class turns out to be not really ready for the following class. In view of such situations, it is desirable to predict the student's success in the following class as accurately as possible.

How student's success is predicted now. At present, the student's success is predicted based on the student's grade in the previous class.

In each class, there are usually several occasions on which the student's knowledge is gauged: exams, quizzes, labs, projects, homework assignments, etc. The grade g for the class is usually a weighted combination of grades $g_1, \ldots, g_n$ for all these instruments for gauging the student's level of knowledge:

$$g = w_0 + \sum_{i=1}^{n} w_i \cdot g_i,$$

with appropriate weights w_i.

How can we improve the current way of predicting the student's success. The weights assigned to different instruments are usually selected based on the instructor's subjective understanding of the importance of different topics.

Because of this subjectivity, these weights are not always a very accurate representation of the material's important:

- the material that the instructor believes to be very important for the following class may be not, in practice, not that important, and

- vice versa, the material that the instructor believes to be not very critical may turn out to be more important than the instructor thinks.

Thus, to get a more accurate prediction of student's success, it is desirable to replace the *subjective* weights with the more *objective* weights – weights for which the corresponding weighted combination is the best fit for the grade in the next class. In other words, once we know, for sufficiently many students $k = 1, \ldots, K$, their grades $g_i^{(k)}$ for different assignments i from the first class and their grades $s^{(k)}$ for the second class, we can use, e.g., the Least Squares techniques to find the values w_i of the weights which provide the best fit for the following equalities:

$$s^{(k)} \approx a_0 + \sum_{i=1}^{n} a_i \cdot g_i^{(k)}.$$

This idea was proposed and analyzed in [2].

What we do in this paper. Our original idea was to simply apply the above general technique a specific example of first two classes from the Computer Science introductory sequence. We expected a routine application, but what we found out was rather unexpected: that to adequately predict the student's success in the following class, we need to go beyond reliable grades. This surprising result is presented in this paper.

2 Case Study: Description

General description of the situation: CS1 followed by CS2. For Computer Science students in the University of Texas at El Paso Computer Science program, the first computer science class "Introduction to Computer Science" (CS1) is a pre-requisite for the next class "Elementary Data Structures and Algorithms" (CS2). This sequence is in line with the 2013 Computer Science Curriculum [1] approved by the Association for Computing Machinery (ACM), the main Computer Science organization.

As a test case, we considered students who successfully took CS1 in Fall 2014 and then took CS2 in Spring 2015.

How the knowledge of CS1 students was gauged. The knowledge of CS1 students was gauged by three midterm exams, a final exam, and 13 labs. The overall grade is a weighted combination of the grades for the exams and of the grades for the labs. To pass the class, the students must gain at least 70 points out of 100.

Midterm exams and final exams are performed in-class. All exams are proctored, so we are confident that the results of each exam properly reflect the students' knowledge. Since the exam time is limited, problems presented at an exam have to be reasonably simple, to enable students to successfully solve them during the time allocated for the exam.

In contrast, labs are performed by students on their own time. Usually, a lab is due a week after it is assigned. Each student's work is supposed to reflect this student's individual work. Students are prohibited from seeking help with working on the lab. However, while the students are not allowed to explicitly ask for *specific* help with the lab assignments, they are encouraged to ask instructors and teaching assistants (and more advanced students from the class) for *general* help with understanding the material and with solving similar problems. Since students get help (indirect but still help) while working on the labs, the grade on each lab does not necessarily adequately reflect the student's ability to individually solve the corresponding problem: in our experience, there is no guarantee that without outside help students will be as successful in solving a similar problem.

Since we wanted the overall grade to reflect the student's individual knowledge, we were hesitant to give much weight to the labs when computing the overall grade for CS1. As a result, each lab was worth only 2 points out of 100, so that overall, the grade for all 13 labs could contribute, at best, to 26 points out of 100. An additional reason not to assign too many points for the labs is that labs are not proctored, so assigning too many points for the labs would create a temptation for cheating. With the current 26 points assignment, even if a student cheats on all the labs, this student still needs to show reasonably good knowledge on other assignments to gain 70 points needed to pass the class.

How the knowledge of CS students was gauged. Two sections of CS2 were taught by two different instructors. While the two instructors agreed on what level of knowledge corresponds to passing the class, they used different weights to assign grades for individual assignments and different thresholds for passing. To compensate for this difference, we multiplied the grade of the first instructor by 1.07 and the grade of the second instructor by 1.03. This way, the passing grade of both instructors becomes equal to the same 70 points threshold as for CS1.

Resulting grades. The resulting grades – sorted in the decreasing order by the re-scaled CS2 grade – are presented in the following table. In this table:

- e_i is the grade for the i-th midterm exam (out of 100),
- ℓ is the grade for the labs (out of 26),
- e_f is the grade for the final exam of CS1, and
- s is the (re-scaled) grade for the second class (CS2), also calculated out of 100.

Comment. Please note that some exams and labs include extra point questions, so some students got more than 100 points.

e_1	e_2	e_3	ℓ	e_f	s
96	108	87	25	102	112
95	100	95	27	102	107
91	101	95	27	109	106
93	102	75	26	107	106
89	107	117	27	103	103
93	106	100	26	102	100
90	0	90	26	86	96
79	99	92	26	98	93
96	97	113	27	106	93
85	99	98	25	82	92
93	98	98	24	94	91
89	100	78	25	98	91
76	92	85	25	89	90
86	99	107	27	10	90
83	80	103	25	94	89
90	101	83	27	95	88
85	106	103	27	101	87
94	97	92	25	88	85
98	98	100	26	101	85
86	101	60	27	51	82
89	97	83	26	105	82
85	82	92	23	95	81
84	96	85	26	98	81
81	93	65	24	94	79
90	84	85	26	99	79
90	98	87	26	76	78
86	76	77	24	84	77
88	91	92	25	86	77
81	97	85	25	82	75
51	87	90	23	92	71
90	101	83	23	81	71
85	107	115	22	84	70
96	83	80	25	99	68
80	81	68	20	70	68
94	92	88	20	76	64
96	83	87	24	108	64
83	92	92	20	82	59
85	87	68	26	88	49
89	52	57	18	92	46
96	97	88	18	89	37

3 Case Study: Analysis and Its Results

Least square regression. We used least squares to come up with a linear function that provides the best fit for the above data:

$$s = a_0 + a_1 \cdot e_1 + a_2 \cdot e_2 + a_3 \cdot e_3 + a_\ell \cdot \ell + a_f \cdot e_f.$$

As a result, we got the following values:

a_0	a_1	a_2	a_3	a_ℓ	a_f
-62.02	0.05	0.01	0.20	4.36	0.14

We see that the coefficient at the lab grade ℓ is much larger than the coefficient at the exam grades, which means that the grade on the labs is a much more important predictor of the success in the next class than the grade on the previous class's exams.

To make a fair comparison, let us re-scale the lab grade ℓ, i.e., replace the original grade ℓ whose maximal value is 26 with a re-scaled lab grade $g_\ell = \dfrac{100}{26} \cdot \ell$ whose range is from 0 to 100 (the same as for each of the exams). Then, the coefficients of the re-scaled linear regression

$$s = a_0 + a_1 \cdot e_1 + a_2 \cdot e_2 + a_3 \cdot e_3 + a'_\ell \cdot g_\ell + a_f \cdot e_f$$

take the following values:

a_0	a_1	a_2	a_3	a'_ℓ	a_f
-62.02	0.05	0.01	0.20	1.13	0.14

The coefficient $a'_\ell = 1.13$ at the re-scaled lab grade g_ℓ is more than 5 times larger than the largest coefficient $a_3 = 0.20$ at the exam grade. In this sense, we can say that the lab grade is at least 5 times more important to predict the grade in the next class than the grades on the previous class's exams.

Conclusion. At the end of CS1, we have:

- grades for the exams which provide a very reliable knowledge of the students' ability – but only about the students' ability to solve simple problems;
- grades for the labs, which gauge the students' ability to solve more complex problems – but much less reliably.

It would have been nice to be able to predict a student's success in the next class based on this student's reliable grades (i.e., grades for the exams). Unfortunately, our analysis shows that this is not possible: to get a good prediction, we need to go beyond reliable grades and take into account lab grades – which gauge the student's ability to solve complex problems but which are not as reliable as the exam grades.

Acknowledgments. The authors are thankful to the faculty of Computer Science department for providing their data, their advise, and their encouragement.

References

[1] *Computer Science Curricula 2013: Curriculum Guidelines for Undergraduate Degree Programs in Computer Science*, ACM Press, New York, 2013.

[2] S. Niwitpong, M. Chiangpradit, and O. Kosheleva, "Towards optimal allocation of points to different assignments and tests", *Journal of Uncertain Systems*, 2010, Vol. 4, No. 4, pp. 291–295.