



A Method for Normalizing Students' Scores

Case Study Faculty of Information Technology

At Al-Ahliyya Amman University

Khaled Al-Jubori , Mustafa Yaseen,

Faculty of Information Technology, AL-Ahliyya Amman University, Amman, Jordan

Abstract

Assessment and evaluation of students in a percentage based grade system is proven to be unfair, biased and in many cases lead to unrealistic evaluations, and unreasonable results. This is basically due to lack of understanding by instructors, who act just as a calculator without any considerations to any measures or rules; i.e. they just add up to totals of exams and assignments given to their students, no adjustments are done, and when done, they are not fair. Unfairness comes from adjusting all grades by an x amount of marks that are either added (to raise the grades) or deducted (to lower the grades). In this paper we are proposing a system that is based on a mathematical approach for adjustments of grades based on criteria that we'll prove to be sound, efficient and fair to all students in any course at Amman University, with an instructor based on their performance and compared to the rest of the students in the specific class.

Keywords: polynomial, instructor, grade, parameter

1. Introduction

Evaluating and grading students is one of the most challenging aspects of teaching. It is not an easy task to design a test that provides students with different learning style and an opportunity to demonstrate their understanding the material, oftentimes; we decide to assess what is most readily assessable because it is difficult to imagine how we might conduct more authentic types of assessment. As with other aspects of teaching, evaluation practices and processes need to be shaped by course goals and objectives.

The thinking, skills, applications, or operations that are the most important manifestations of learning in a course should be the basis for evaluation and grading.

In order for a student to learn from and value your feedback, it must be shared before they have moved on to other work, providing a quick response to students it is challenge particularly for faculty of information technology with heavy student load. We realize that most frequent concerns of Information system and computer science students is how fair they will be evaluated student feel very insecure when they cannot predict final score, their complains that the type of row percentile marks we are using at Al Ahlyia Amman University is unfair.

As outlined by many experts, instructor's identification of goals and objectives for courses must be followed closely by the design of evaluation of the grade system. We know that evaluations of

students learning and the assigned marks is every one's interest to try to make the evaluation system fair and free from irrelevant errors as possible. Thus, any activity used to evaluate students' learning outcome must be an accurate reflection of the skill or concept under testing for evaluation; this means that Evaluation activity must appear to be evaluated to the course content. A common complaint of our students is that tests are not related to the content of the course, or to what was presented in class. Although we recognize that the things we assign directly related to the course, the students often don't get the connection (Deans notes reports to Instructors), and student's surveys.

- Evaluation activity of the performance predicts performance on other closely noted skills, either at the same time (current validity) or in the future (predictive validity), that is to say, the test is highly correlated or not. one of the main reason the instructions is using

slide show materials from the internet sources while the text book in our collage has quite different approach

- Any activity used to evaluate students' learning outcome must be reliable, implying that the system will produce the same evaluation of performance each time. Here the goal is to eliminate as many sources of errors as possible and accept the fact that errors will occur any way.
- Some of the major sources of errors in evaluating students may occur due to poor communication of expectations that is student's failure to correctly interpret the problem, and lack of sufficient information about performance this is our main problem with new instructors. Thus one of the most important steps in selecting an evaluation is to analyze the objectives and design activities aimed directly at the content and level of those objectives. This way we can reduce the errors in students' evaluation.
- Any activity used to evaluate students must be recognizable, students should be aware of how they will be evaluated and their class activities should prepare them for those evaluations. Exams should not be a game of guess. One of the biggest complaints of students is that the basis for evaluation was unclear to them. In brief, an instructor should choose evaluation types/styles which are clearly related to the content and class activities of the course. Tests should be a mirror of teaching; it should never be the case that students come to the exam not knowing what to expect. Students don't mind hard exams as long as there are no surprises and they can recognize the relationships between the exam and the course topics. here its again instructor's attitudes is only to penalize the students by their own exams.
- Any activity used to evaluate students should be realistic, that is the amount of information obtained is balanced with the amount of work required. Often, instructors forget that our students are taking 5-6 courses per semester, they also forget that they (themselves) are teaching 2-4 courses per semester, so as much as instructors he must face the overestimated heavy work in which neither the student instructor can bare that is the whole pr collapse under its own weight.



$$y_i(x_i) = (\mu + 1)x_i - 0.01\mu x_i^2 \quad (2)$$

3. The Method Properties and Perceived Fairness

This method will use almost raises student's grades, and the course average for the adjusted score will be higher while the standard deviation is almost sustain its normal value, this will result in same distribution, and that an evidence of no flow of adjustment method. On the other hand, if the material is being tested is something critical like Object Oriented programming course where overall class performance is very poor, and the instructor himself don't mind but he doesn't have the scheme to adopt , or presently don't want to curve the grades randomly, at the same time there are many students complains about un fair grading against him , our formula to correcting the course grades perceived deficiencies without watering it down may be made with many corrections adapters .

On the other hand, a survey of statistical analysis at Amman university (2008-2009), table (1) shows that the main problem lies of the students having average marks below 60 and this means more than 28% of our students are under university regulation critical warning, the big challenge to the University, is how to deal with this serious problem finding proper and fair solutions is the major problem.

Number of students	Percentage	Frequency
336	5.3%	less than 50
426	6.8%	55-50
1017	16.1%	60-56
1389	22%	65-61
1069	16.9%	70-66
868	8.13%	75-71
569	9%	80-76
355	5.6%	85-81
209	3.3%	90-86
66	1%	95-91
6	1%	100-96

Table (1)

4. Application:

The class in Table (2) and figure (1) illustrates some of the problems; an instructor faces in the odd scores of his 54 Object Oriented Programming students class at our faculty. This class has a lower average score compared to all other classes at IT, the figure shows slightly skewed distribution with a few more students clustered near the lower end of all intervals <60, this means that 57% of the students in the class has stragglers score below 60 (this scores is much lower than everyone else), and according to the university rule those students receive an academic warning.

From the Table (3) and Figure (2), and after we apply the equation (2) to the students' row scores, we can see the effects of these various adapting parameter μ on the distribution of the grades

Also we notice, from the table(4) below ,and at the various grading parameter μ , the results are almost have same standard deviation but all average up which is of course an advantage to the educational processes ,at $\mu = 0.4$,we have produced shaped curve of grades , see Table(3) and Fig (2) www.IJCSI.org average student score = 66, which is around the faculty average ,note that the low score marks < 60 , which is our main focus is

Because of the marks policy that we adopt at Amman University is so closely tied to personality of teaching which means view our own roles as teachers and it is so obvious from the analysis made from last semester statistical analysis ⁽¹⁰⁾ to our final row scores which showed a gap between instructor's evaluations row policy (neither absolute performance level nor relative performance level).

If we set our evaluation criterion as a performance measure, then the score reflecting poor performance should be dropped, and this is very dangerous, however in order to make such an evaluation, the exams need to be correlated into common evaluation system, specifically they need to be placed upon standard scale for comparison.

Therefore, using row scores to evaluate final grades may not accurately capture a student's true performance within a class. This point is a challenge to any education institute and in particular to policy makers.

We would like the distribution of individual student's performance for all exams to be fair, despite differences in time, instructors, teaching different course and any other factors. Only then can evaluations be considered comparable. Without this common scale error in grade will results. and then we have to bear the consequences. and here what we are trying to do.

2. Formula for computing the adjusted grads Total

The scores will be based on a national percentile ranking. A score at 50th percentile is by definition the low passing grade. Because the class Average in FIT (Report No2) is generally somewhat higher than 50%, we must adjust the row percentile score in order to reflect the grading standards that we use. Now the class average on final examinations on FIT courses, as an example we'll take Second Term in academic year 2007-2008; is around (62%). Therefore, scoring at the 50th percentile on any exam should receive a grade of 62/100 on the final to count towards any course grade. The challenge then is how to convert this type of analysis into a general formula. Whatever method of conversion is used, it must be fair, and it must give as artificially higher boost to students who score at the very top or the very bottom of the percentile markings. To compute the student adjusted total points, from the row score measure

Consider the nth degree polynomial:

$$y_i(x_i) = \alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2 + \dots + \alpha_n x_i^n \quad (1)$$

If we work on this polynomial, to insure the method is applicable to adjust grades to normality we must impose the following conditions:

- $\alpha_0 = 0$, that is there is no constant increase to any student's marks in the class.
- We will take three terms of polynomial, since at the end we are dealing with input- output positive integer scoring points only and no decimal digits is allowed.
- For each class section there is an adjustment parameter μ varies from zero to maximum one depending on the class average of the student in each class
- Clearly minimum scoring point is zero and maximum is hundred.
- The adjusted point total can never be lower than the original point total; this condition avoids an innumerable conversation with students where one tries that is a Panellizing student is not allowed.
- x_i is the row score of the students, and y_i is the final adjusted score.

Now imposing those above conditions, we simply conclude the adjusted second-degree clear polynomial:

dropped from 57% to 33%, which is very good percentage and it is in our university range, over all this is the best we can do to the students' scores.

Intervals	Frequency	%
0-49	18	33
50-59	13	24
60-67	5	9
68-75	3	6
76-84	6	11
85-100	9	17
Total	54	100

Table (2)
 $\mu = 0$, Row Average Marks=57

Intervals	Frequency	CrossMark 9% for update
0-49	12	22
50-59	6	11
60-67	8	15
68-75	10	19
76-84	12	22
85-100	6	11
Total	54	100

Table (3)
 $\mu = 0.4$ Average Marks=66

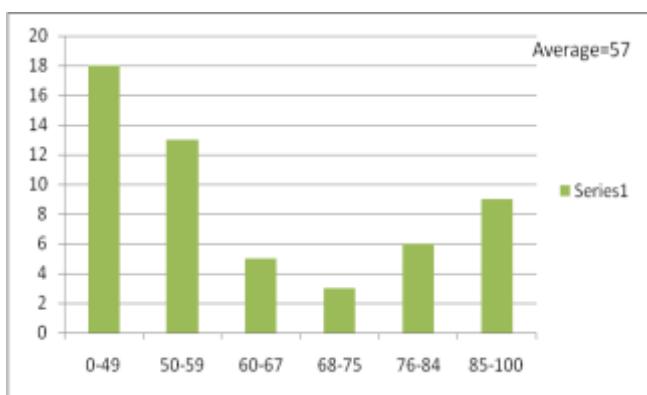


Figure (1)

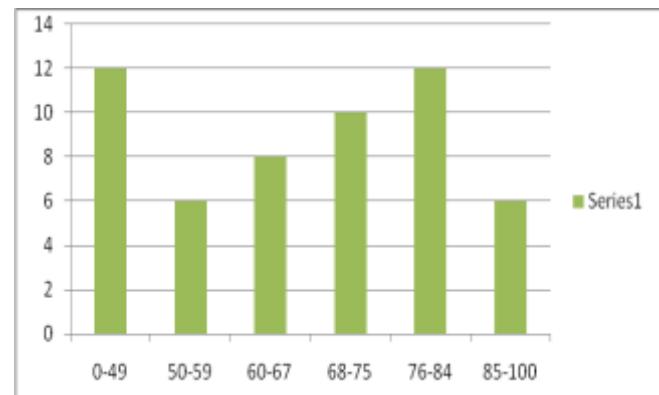


Figure (2)

Intervals	Row Scores	$\mu = 1$	$\mu = 0.9$	$\mu = 0.8$
0-49	18	0	0	0
50-59	13	4	9	12
60-67	5	10	8	6
68-75	8	7	7	8
76-84	10	11	9	10
85-100	0	22	21	18
No. of students	54	54	54	54
Average	57	79	77	75
STDEV	16	13	14	14
Total sum	3102	4291	4172	4053
Total Increase	0	1070	951	832
Average increase per student	0	20	18	15

Table (4)

5. Summary and Conclusions

We introduced a simple and fair formula to our student's odd scoring problem, and with some other normalization schemes we came out with this proper method to deal with our students grading problem in which the instructor can use these adjusted scores with assurance of fair grading. We use this adjusted method on several different courses, we note that the adjustment is straightforward to compute and has several desirable properties, no draw back at all, and surprisingly the outcome results is acceptable by the teaching and the whole students.

$\mu = 0.7$	$\mu = 0.6$	$\mu = 0.5$	$\mu = 0.4$	$\mu = 0.3$	$\mu = 0.1$	$\mu = 0$
0	3	6	12	14	18	18
14	12	12	6	9	8	13
4	8	8	8	9	9	5
12	10	7	10	5	6	8
7	6	11	12	13	13	10
17	15	10	6	4	0	0
54	54	54	54	54	54	54
73	71	68	66	64	60	57
14	14	15	15	15	15	16
3934	3815	3696	3577	3459	3221	3102
713	594	475	357	238	119	0
13	11	9	7	4	2	0

It is a challenge to deal with abnormal grading results, and while you receive a lot of complains from majority of student's affair normalized schemes comes and at the end you will have affair. Practical presentations show that this method is well designed and judiciously applied and reliable, easy to implement and explain, perceived to be fair tool and we can use it to reduce the effect of inconsistency of different scoring measures.



Acknowledgements

We developed the grading adjustment program [10] jointly with our research assistant Abed Alquisi we extend our thanks to him who proved to be comments on the program.

References

- [1] Tunick D and Schechter D; "Normalization of First Year Law School Grades". College and University, v58 n2 p159-81 Win 1983.
- [2] Jacobson N; "A method for normalizing students' scores when employing multiple graders", ACM SIGCSE Bulletin, Volume 33 , Issue 4 (December 2001)
- [3] Johnston I; Herrmann S; and Conklin J. "A Model for the Analysis of Inflated Grades". Eastern Oregon University Science Journal, (1998).
- [4] "Grading Practices," at uga.berkeley.edu/sled/bgd/grading.html, an update of Chapter 32 of Davis, Barbara Gross. (1993). Tools for Teaching, San Francisco: Jossey-Bass Publishers.
- [5] Shub C, Astrachan O, Levine D, Reges S, and Walker F; "Faster, fairer, and more consistent grading, using techniques from the Advanced Placement reading" (panel session), Proceedings of the twenty-first SIGCSE technical symposium on Computer science education, p.266, February 1990, Washington, D.C., United States.
- [6] "Normalizing Section Grades: Barnes-Billingsley contraction formula" at Hebel S, "Mikyung Ryu: Grades the states on higher education" Chronicle of Higher Education, 7/15/2005, Vol. 51 Issue 45, pA15-A15
- [7] Cliffordson C, "Differential Prediction of Study Success Across Academic Programs in the Swedish Context: The Validity of Grades and Tests as Selection Instruments for Higher Education. Educational Assessment; 2008, Vol. 13 Issue 1, p56-75
- [8] Gispert JD, Pascau J, Reig S, Martinez-Lazaro R, Molina V, Garcia-Barreno P, and Desco M, "Influence of the normalization template on the outcome of statistical parametric mapping of PET scans. Neuro Image;Jul2003, Vol. 19 Issue 3, p601, 12p
- [9] Neil A. Weiss, Introductory Statistics, 6th edition, Addison Wesley, 2002
- [10] MSNCONVERSION- Package