Evaluation Reforms: Moving From Absolute to Relative Grading with Q-Factor Control

[1]Suren Patwardhan, [2]Rachana Desai
[1,2]K J Somaiya School of Engineering, Somaiya Vidyavihar University
[1] surenpatwardhan@somaiya.edu
[2] rachanadesai@somaiya.edu

Abstract*— In the evaluation reforms, a key component is the changeover from absolute to relative grading system. In the learner-centric education paradigm, the relative grading offers certain advantages. While many premier institutes tend to follow a relative grading system, it is important to check the applicability of this shift in the context of institute-specific requirements and controllable parameters for any institute planning to change from absolute to relative grading. In this article, analyses of grade distribution and comparison of evaluation methods – the absolute grading and four types of relative grading is presented using two data sets. Marks scored by a group of 300 students in two courses where, one is a high-scoring subject and the other is a low-scoring subject are considered. The four types of relative grading systems presented here are scale-up: with and without scaling factor, percentile method used in most of the competitive examinations and the normal distribution method using the median and standard deviation. All the graphs are converted to bell-curves for uniformity in comparison. The analyses shows when one grading method is suitable for a particular course, it does not yield similar result with some other course/s. While each system including the absolute grading has its own advantages, the appropriate choice of a particular system and its success depends upon the flexibility it offers. In this article, the normal distribution method is further modified to add a quality factor (Q-factor) to maintain a minimum level of competency that the students are expected to attain. With this modification, it is showed that this modified normal distribution method of relative grading can be tailor-made to diversify evaluation and grading standards. The implementation of a particular system depends upon the requirements of the institute, its attributes, effectiveness and productivity of examination system and hence these models need to be customized accordingly.

***Keywords*—absolute grading; bell curve; evaluation reforms; normal distribution; relative grading**

ICTIEE Track*: Assessment of Effective Teaching*
ICTIEE Sub-Track: Diversifying assessment approaches: Success stories

I. INTRODUCTION

FOR any educational system, the curriculum revision is a periodic and necessary process in order to tune-in with the changing times (P. S. Aithal & P. M. Suresh Kumar 2016). Evaluation reform may come as a part of curriculum revision or as a separate process. For the latter case, some of the key influencing parameters are: effectiveness and productivity of examination (Sandeep Kumar Sharma 2014), provide development-inducing feedback for efficiency and effectiveness of the system (Sharon Campbell-Phillips 2020), enhancement of graduate attributes (Aithal P. S., & Suresh Kumar P. M 2015), quality enhancement in higher education institutes (Aithal P. S., Srinivas Rao A., & Suresh Kumar P. M. 2015) and entry of blended learning in education (Randy Garrison D. & Heather Kanuka 2004). There are some global common observations on traditional evaluation. They include too much emphasis on memory-recall, lack of testing higher-level cognitive skills such as logical reasoning, critical thinking, comprehension, emotional intelligence etc., fate of students decided by single vertical of assessment (final examination) and so on (UGC 2019). The University Grants Commission (UGC) in its National Education Policy (NEP) 2020 suggests collaborative and peer learning in higher education to ensure every student's learning. In evaluation reforms, this is often addressed by adopting relative grading system since it is regarded as norm-referenced evaluation system (Sang-Keun Shin 2018). The key purpose of evaluation reforms includes education for overall development of the student, to develop learning outcome-based education (LOBE) and to create multiple horizontal assessments (Mohd. Beyyoudh et.al. 2019).

Although the terms Assessment, Evaluation and Examination are sometimes used interchangeably, there are some fundamental differences between them. Assessment is usually process-oriented while evaluation is product-oriented. Assessment tests how something is learnt while evaluation tests what is learnt. Assessment is generally formative type. It monitors students during instructions whereas evaluation is generally diagnostic. It provides feedback on what can be improved (Sandeep Kumar Sharma 2014). Yet, both are evidence-based meaning a concept or strategy derived from objective evidences e.g. a mind map made by student or marks obtained in an examination (R. V. Mahendra Gowda & N. Viswanathan 2022). The examination in itself can be defined as "a quantitative measure of learners' performance and is usually held at the end of the academic session or semester" as per University Grants Commission (UGC 2019).

Evaluation reforms are often connected with a shift from

Suren Patwardhan,
K J Somaiya School of Engineering, Somaiya Vidyavihar University.
surenpatwardhan@isomaiya.edu

traditional absolute grading to relative grading. Both the methods have certain advantages and disadvantages and it is important to systematically analyze the impact of a particular grading method.

Several articles have been published on absolute and relative grading system discussing their comparative analysis for implementing in engineering institutes (R. V. Mahendra Gowda & N. Viswanathan 2022), effect of relative grading on passing standards (Ayfer Sayin 2016) and relative grading as equitable approach for students' evaluation (Johanson, George A. 1993). (Jon J. Denton 1989) discussed factors influencing the decision of selecting appropriate grading system. The absolute and relative grading standards are discussed in details by (George R. Quinn & Charles A. Szuberla 1963). Difference between absolute and relative grading can be done using various parameters. Table I lists these parameters with brief meaning of each for the respective grading methods.

TABLE I
ATTRIBUTES OF ABSOLUTE AND RELATIVE GRADING

Parameter	Absolute Grading	Relative Grading
Method	Criteria-referenced	Norm-referenced
Description	Grades based on student's performance w.r.t a pre-defined standard (e.g. 40 marks passing)	Grades based on student's performance relative to his/her peers
Flexibility	Rigid. Marks obtained are marks obtained	Marks obtained can be normalized or scaled up/down
Sensitivity	Insensitive to difficulty level of subject, stringent evaluation etc.	Can be tailor-made depending upon level of difficulty etc.
Customizability	Number of unsuccessful students is fixed	Number of unsuccessful students can be varied
Sample size	Works on both, small and large sample size	Doesn't work effectively on small sample size (typically below 30) (R. V. Mahendra Gowda & N. Viswanathan 2022)
Applicability	Necessary when students need to achieve a certain skill set or certain established level of competency	Useful when attainment of certain level of competency may not be important but performance w.r.t. peers is sufficient

II. Scope of Study

In engineering curriculum, some courses show high-scoring trends while some courses consistently show low-scoring trends. There are multiple factors for such low-scoring trend for some courses such as difficulty level of the course, difficulty level of the question paper, level of assessment etc. It may not be possible to address each and every factor influencing the overall low or high grades obtained for a particular course. However, in this paper, the author shows that incorporating a quality factor to the normal distribution method of relative grading, a tailor-made approach can be adapted to address most of the factors influencing students' performance for justifying assessment of effective teaching. Analysis of existing grading methods and issues associated with them such as minimum passing grades or cut-off mark for highest grades is done for making evaluation system more applicable to different conditions (Nagaraju Dasari 2021). While the usefulness of relative grading or norm-specific assessment is a subject of discussion since many years (George R. Quinn & Charles A. Szuberla 1963), studies have also shown importance of absolute or criteria-based assessment while adopting outcome-based approach in education (Beatrice Lok et. al. 2015). With the vivid nature of requirement for different streams and types of courses, some studies suggested to go beyond the dichotomy of relative or absolute grading and adopt a complementary approach for proper customization and effective evaluation (Beatrice Lok et. al. 2015), (Liz Freeman & Andy Miller 2001), (George R. Quinn & Charles A. Szuberla 1963), (Sang-Keun Shin 2018). While most of the articles mentioning complimentary approach suggest switching between systems based on the course and learning requirements, a radically new approach to directly incorporate the criteria-based benefits of an absolute grading system into the framework of norm-based relative grading system is missing. In this article, authors tried to bridge this connection by introducing an additional control parameter (defined as quality factor in the next session). The articles discuss impact of this factor in re-distributing the grade table and while maintaining the evaluation standards.

TABLE II
GRADE TABLE – KJSCE, SOMAIYA VIDYAVIHAR UNIVERSITY

Range	Grade	Pointer	Remarks
80 & Above	O	10	Outstanding
70-79.99	A+	9	Excellent
60-69.99	A	8	Very Good
55-59.99	B+	7	Good
50-54.99	B	6	Above Average
45-49.99	C	5	Average
40-44.99	D	4	Pass
Below 40	F	0	Unsuccessful

III. Data and Methodology

The final score in any course in KJSSE consists of performance of students in the end-semester theory examination (ESE) and continuous assessment (CA), totaling to 100. KJSSE has implemented relative grading system from academic year 2023-24. The marks obtained out of 100 over an absolute scale by all students are used to generate grades by using relative grading method, which, in this article is referred to as "restricted scale-up" method. It is explained in the next subsection. The relative score obtained by this method is used to generate grades as per Table II. For every letter grade such as O, A+ etc., a pointer is defined with pointer 10 being the topmost while a pointer below 4 is regarded as fail. This pointer for a particular course along with number of credits for that course is then used to find semester grade performance index (SGPI).

Final examination scores of 300 students of First Year Bachelor of Technology of mixed branches of Engineering such as Information Technology, Robotics and Artificial Intelligence, Mechanical Engineering etc. from K J Somaiya School of Engineering (KJSSE) was collected. KJSSE is a part of Somaiya Vidyavihar University (SVU) which is a Mumbai based state private university.

For comparative analysis, marks scored by same students in two different courses are considered. Here these courses are referred as course 1 (C1) and course 2 (C2). In C1, 23% of the students (69 in all) got topmost grade in the absolute score with the highest score of 97 out of 100. In C2, none of the students got any of the top two grades in the absolute score with the highest score of 68 out of 100. As it can be easily learnt, C1 is very high scoring course while C2 is a low-scoring course. By choosing such courses with extreme performances by students, a wide range of scalability can be tested for the various relative grading methods discussed in this article. Following subsections section give description of the type of relative grading methods considered in this article along with their corresponding formulae:

A. Absolute grading

This is the simplest method where the scores in the examination are used as it is without any scaling factor or modification. For generating bell curve, the pointers generated as per the grade table II are used by applying the standard formula $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{(x-\mu)}{\sigma}\right)^2}$, where σ is standard deviation, μ is mean value, x is the i^{th} entry and f(x) is probability density factor for i^{th} entry.

B. Relative grading

All relative grading methods are based on peer performance. The scores obtained by students are compared against a specified formula and an effective distribution is obtained, which gives mutual performance of students in a selected data set. Following is a description of the four types of relative grading methods used in this study:

1) Simple scale-up:

This is the simplest relative grading method. In this, the highest score is equated to 100 and accordingly, all the scores are scaled up by an equal multiplying factor.

2) Restricted scale-up:

Here, the multiplying factor is limited to a pre-defined value. This method is currently used in KJSSE. The main purpose of this restriction is to avoid arbitrary inflation of grades and to maintain a certain quality of results.

3) Percentile system:

This type of relative grading is widely used in mostly of the competitive examinations such as Joint Entrance Examination (JEE) or state-level Common Entrance Test (CET). This is similar to scale-up method but it incorporates the number of data values in the data set to obtain the percentile rank of a particular data point. Common method to calculate the percentile is $P = \frac{n}{N} \times 100$, where n is number of data points below the particular data point x and N is the total number of data points and P is the percentile score

4) Normal distribution method:

This type of relative grading is also widely used many universities and institutes such as Indian Institute of Science Education and Research. In this, median score of the data set is used to decide the cut-off /passing marks. Typically, if "M" is the median then the cut-off/passing marks are set at M/2. A control parameter "Sa" is defined such that Sa ≤ highest score in the data set. The interval M/2 to Sa is then (nearly) equally divided into several intervals corresponding to different grades. With respect to Table II, this range will cover six grades from O to D while for any score falling below M/2, grade F i.e. unsuccessful is allocated

IV. Finding/Analysis and Discussion

For the analysis, four different types of relative grading are discussed and compared vis-à-vis absolute marks obtained. The relative grading presented here are: simple scale-up, restricted scale-up, percentile method and normal distribution method. The normal distribution method is further modified to incorporate the minimum desired skill level by using a quality factor or simply Q-factor in this article. All the data sets – absolute and relative are converted to bell-curves using normal distribution formula and normalized to 100% for uniformity in comparison.

Fig. 1 and Fig. 2 show the grades distribution of both courses C1 and C2 using the methods listed in section III. For C1, none of the relative grading yields any significantly differing distribution from the absolute grading. The simple and restricted scale-up grade distributions match perfectly. This depends upon the choice of scaling factor, which is restricted to 1.25 in this analysis. These two methods would start differing when the highest score drops down below 80 in this case. Hence for any course for which, as long as the highest score is above 80, the restricted scale-up method has no potential advantage over absolute or simple scale-up methods for enhancing grade distribution.

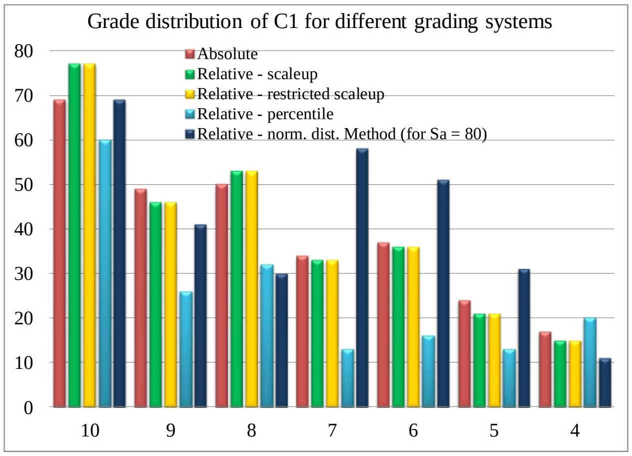

Fig. 1. Number of students per grade for course C1 for different grading systems

Note: In all figures with bar charts, the Y-axis indicates number of students and X-axis indicates grade pointers from 10-4. In all figures showing bell curves, the Y-axis indicates normal distribution value for the specific data point normalized over a scale of 0-100%

For C2 however, the advantage of relative grading scale-up methods is apparent. While the absolute grading showing no student in the top two grades due to low value of the highest score, the scaling factor in the relative grading can be selected based on the difficulty level of course, question paper or assessment. Thus, the scale-up method, even with a modification of restricting the scaling factor is not applicable to all types of courses with the same flexibility in deciding the grade distribution.

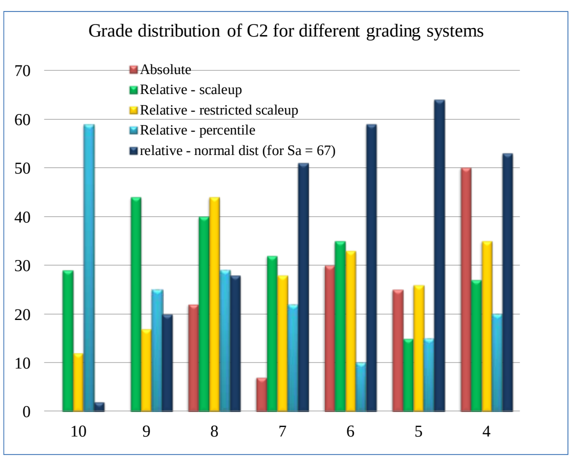

Fig. 2. Number of students per grade for course C2 for different grading systems

The percentile method has no significant effect on grade distribution of C1 with respect to absolute grading but in case of C2, the percentile method generates a large fraction of students with the topmost grade. Such a high fraction of topmost grades may lead to a misleading picture. At the same time, for both the courses C1 and C2, the percentile method has very high number of failures referring to Table III, which can also lead to a drawback when such a large (40%) of the students fail. Moreover, there is no control by which the grade distribution can be modified.

TABLE III
NUMBER OF FAILUERS FOR DIFFERENT GRADING METHODS FOR BOTH COURSES

No. of failures	C1	C2
Absolute	020	166
Relative - scale up	019	078
Relative - restricted scaleup	009	105
Relative - percentile	120	120
Relative - normal dist without Qa-factor	009	023
Relative - normal dist with Qa-factor	065	095

The normal distribution method, on the other hand offers higher flexibility in a sense that number of students securing the topmost grade and overall grade distribution can be tailor-made using the control parameter Sa as defined in the previous section. Fig. 3 and Fig. 4 show the variation of grade distribution with Sa-factor for both courses, C1 and C2. Note that the number of failures is independent of Sa-factor.

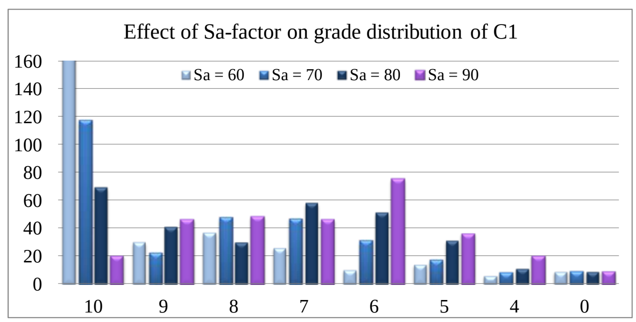

Fig. 3. Variation in number of students per grade for course C1 for different setting of control parameter Sa

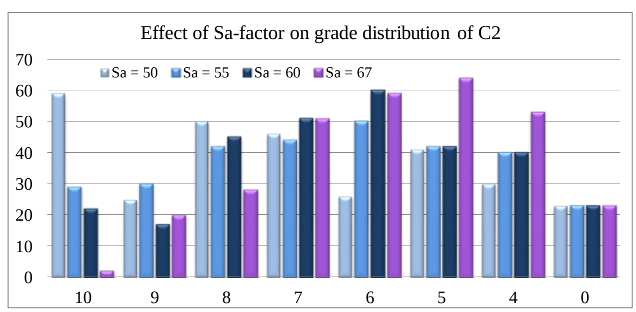

Fig. 4. Variation in number of students per grade for course C2 for different setting of control parameter Sa

Thus by fixing a certain level of passing standard, a feature which is unique in absolute grading can be incorporated in relative grading without losing the flexibility of grades distribution with this normal distribution method.

However, all these relative grading methods - absolute as well as relative - suffer from a major drawback. Following Table IV lists the cut-off marks (marks below which, student is declared fail) for both courses C1 and C2.

TABLE IV
CUT-OFF MARKS FOR DIFFERENT GRADING METHODS FOR BOTH COURSES

cut-off marks	C1	C2
Absolute	40	40
Relative - scale up	40.2	32
Relative - restricted scale up	40.2	28
Relative - percentile	58	34
Relative - normal dist. without Qa-factor	31	18
Relative - normal dist. with Qa-factor	31-52	18-30

We can see that none of the grading methods offer control on the cut-off marks. The cut-off marks can go very low or exceptionally high depending upon the grading method, highest score and score distribution. Traditionally, in any Engineering course, the passing marks are typically set at 40% in the absolute grading. In a typical Engineering course where student securing as high as 58 marks out of 100 may also be declared fail. On the other hand, when students are required to attain a certain level of skill standards, student securing as low as 18 marks out of 100 may also be declared as successful.

Both these extreme examples indicate some kind of discrepancy in the relative grading methods. This can be eliminated by defining a parameter that would control the number of failures in any course based on various criteria of difficulty and standards of passing. In the normal distribution analysis, in addition to the control parameter Sa, another parameter called “quality factor” Qa is introduced. This factor controls the cut-off marks without disturbing the number of students fixed in the topmost grade using Sa-factor. Fig. 5 indicates effect of Sa factor while Fig. 6 indicates the effect of Qa factor on the grade distribution.

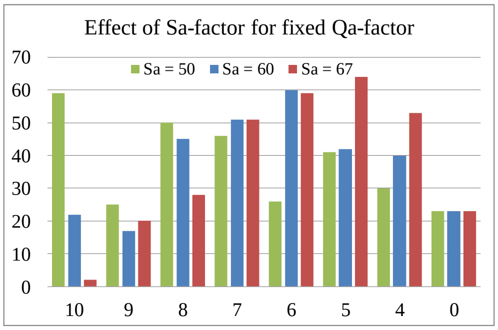

Fig. 5. Variation of grade distribution with Sa-factor for fixed Qa-factor = M/2 = 18 for course C2. Grade 0 indicates number of unsuccessful students

We can see that the Sa-factor, while controlling the number of students in the topmost grade has no effect on the number of failures, the Qa-factor, while controlling the number of failures has no effect on the number of students getting the topmost grade. Thus, Sa and Qa factors can be independently varied over a suitable range to get a complete flexibility of grade distribution. This feature will be very useful while dealing with courses with varied nature (such as theoretical, mathematical etc.) and difficulty level. It is also useful to minimize the variations in assessment of the same course by different assessors.

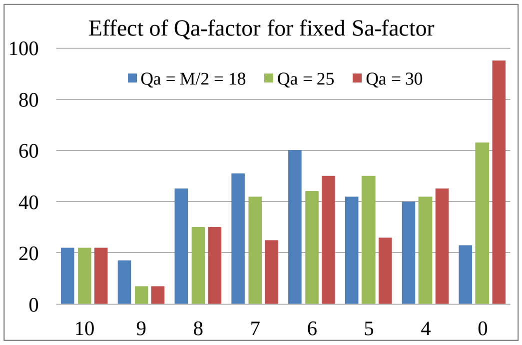

Fig. 6. Variation of grade distribution with Qa-factor for fixed Sa-factor = 60 for course C2. Grade 0 indicates number of unsuccessful students

For a uniform comparison, data sets of all methods including the absolute scores were converted to a normal distribution by using the function defined in previous section. Often, the bell curve generated by a normal distribution is an important indicator of peer performance of students for a course. In addition, the effectiveness of evaluation is also generally judged from such type of bell curve distributions. A symmetric bell curve is indicative of good achievement of objectives set in an outcome-based education system while a skewed bell curve is an indicator of either underestimating or over-estimating the criteria parameters set for measuring missing the objectives of a course. The centre of symmetric bell curve remains fixed for absolute grading system while it is flexible for relative grading system, which is often used by many institutes for grade-redistribution based on the results. Fig. 7 and Fig. 8 show the normalized bell curve for pointers from 10 to 4 corresponding to grades from O to D.

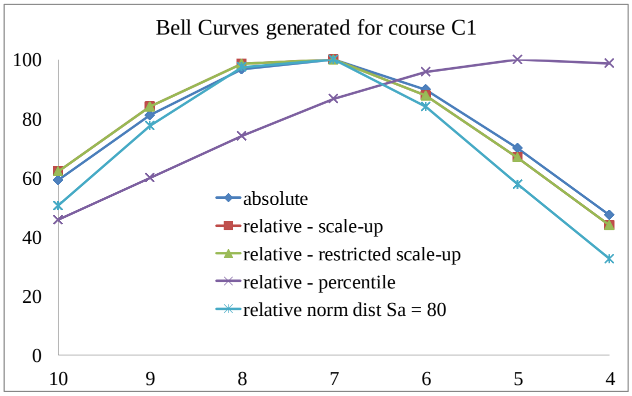

Fig. 7. Normal distribution curve (bell curves) for different grading systems for course C1

It is found that in case of C1, the bell curves of all grading methods - absolute as well as all relative except for the percentile method are almost similar with the peak occurring along pointer 7. The bell curve for percentile method is skewed due to very high cut-off marks. In case of C2, the bell curves significantly differ from each other. Absolute grading method results in a completely skewed distribution for C2.

This can be improved by using the scale-up relative grading. However, this method does not differ from the restricted scale-up relative grading for the case of C1.

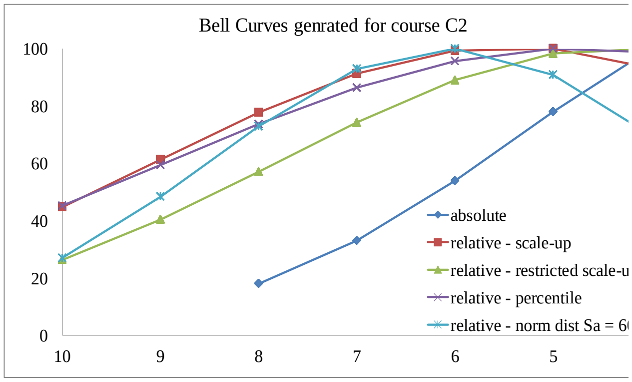

Fig. 8. Normal distribution curve (bell curves) for different grading

In either case, the normal distribution method of relative grading gives much flexibility in terms of adjusting the grade distribution using the Sa-factor as it can be seen from Fig. 9 and Fig.10.

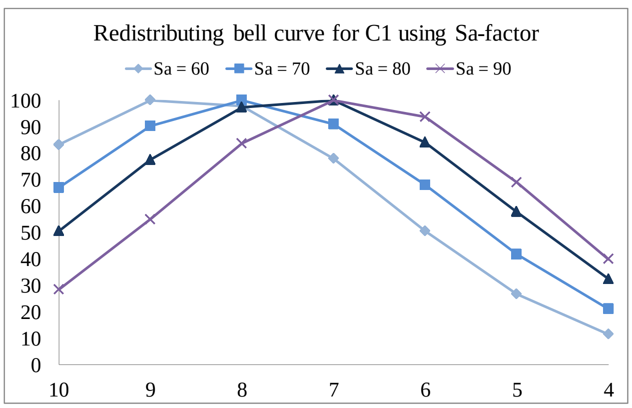

Fig. 9. Adjustment of the Normal distribution curve (bell curves) using Sa-factor for course C1

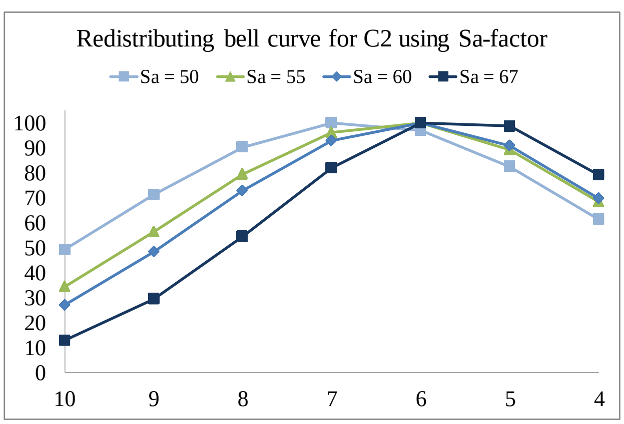

Fig. 10. Adjustment of the Normal distribution curve (bell curves) using Sa-factor for course C2

The Sa factor is solely decided by the assessor and moderator based on different criteria including difficulty level and differences in the assessment. Once a particular value of Sa is fixed for a particular course, further flexibility is achieved by adding the Qa factor to match the pre-defined criterion (such as bare-minimum marks set for a particular course by the instructor). A proper combination of these two parameters by checking the variations within a specified range of these parameters (e.g. 60 to 90 here for Sa factor and M/2 to M for Qa factor) is required for the most effective evaluation of any course.

The implementation of a particular system depends upon the requirements of the institute and hence these models need to be customized accordingly. In this communication, we tried to address this in regards to recent changeover from absolute to relative grading in KJSSE, Somaiya Vidyavihar University.

CONCLUSIONS

Evaluation reforms often incorporate a shift in the grading system from absolute to relative. At the background of modern reforms in education pedagogy in general, the relative grading is proved to be effective in outcome-based education and peer-learning. However, in making a shift from absolute to relative grading, it is also important to capture the advantages of absolute grading system. Absolute grading becomes useful when students are required to attain a specific skill level of capacity. This feature usually gets overlooked in traditional relative grading systems. By suggesting a modification to existing relative grading system, in this article we have tried to show analytically how the skill level requirement can be tailor-made without disturbing the overall grade distribution as well as the number of students assigned the topmost grade. Educational institutes switching to relative grading are suggested to do a comprehensive statistical analysis by the methods presented in this article. The choice of Sa and Qa factors needs to be decided based on nature of course, performance of students (absolute marks), inputs from the course instructor, bare minimum passing standards set and grade distribution fulfilling the objectives of the course. Of all the relative grading system discussed here, authors propose that the normal distribution based system is better adoptable for implementing relative grading system and it can be further improved by adding the Q-factor control for maintaining quality of results.

LIMITATIONS AND FURTHER SCOPE

In this analysis, data sets of two courses with extreme results were considered. For a better analysis, applying the analysis to more permutations of scores as well as comparison of grading methods for formative evaluations in addition to summative may lead to better picture. This analysis provides basic idea of incorporating the advantages of both, absolute and relative grading. A comprehensive analysis based on a variety of courses can be the next possible study.

ACKNOWLEDGEMENTS

The authors are grateful to the management and administration of K J Somaiya School of Engineering, Somaiya Vidyavihar University, Mumbai for their constant support, recommendation and encouragement for research in teaching/learning and evaluation. The authors are thankful to the examination cell of K J Somaiya School of Engineering for providing the data for analysis. Sincere thanks to the team formed for adopting relative grading led by Dr. Pushpendra Rai and members viz. Kartik Patel, Kiran Thale, Poonam Bhogale, Ravindra Salvi, Dr. Sujata Pathak and Dr. Vilas Dhore with constant support from our deputy director Prof. Deepak Sharma. Authors also appreciate valuable inputs given by KJSSE Teaching-Learning Centre (TLC) and thank them for giving opportunity to showcase our work by organizing TLC conclave as well as motivating to write an article in this field.

References

Sandeep Kumar Sharma (2014). Reforms in Examination and Education System: "India has examination system, not education system." *International Journal of Research*, 1(5).

Sharon Campbell-Phillips (2020). Education and Curriculum Reform: The Impact They Have On Learning. *Budapest International Research and Critics in Linguistics and Education Journal*, 3(2), 1074-1082.

Aithal P. S., & Suresh Kumar P. M (2015). Enhancement of Graduate attributes in Higher Education Institutions through Stage Models. *International Journal of Research in Business Management*, 3(3), 121-130.

Aithal P. S., Srinivas Rao A., & Suresh Kumar P. M. (2015). Quality Enhancement in Higher Education Institutions : A case study of SIMS. *International Journal of Multidisciplinary Research and Development*, 2(5), 18-31.

R. V. Mahendra Gowda & N. Viswanathan (2022). A Comparative Analysis of Absolute Grading and Relative Grading of Academic Performance of Learners. *International Journal of Advanced Science and Engineering*. 9(1), 2617-2636

Randy Garrison D. & Heather Kanuka (2004). Blended learning: Uncovering its transformative potential in higher education. *The Internet and Higher Education*, 7(2), 95-105.

Sang-Keun Shin (2018). The Characteristics of Absolute Grading of the College Scholastic Ability Test English Section. *English Teaching*, 73(3)

Mohd. Beyyoudh et.al. (2019). Towards a New Generation of Intelligent Tutoring Systems. *International Journal on Emerging Technologies*, 14(14).

Ayfer Sayin (2016). The Effect of Using Relative and Absolute Criteria to Decide Students' Passing or Failing a Course. *Journal of Education and Training Studies* 4(9).

Valentina Paredes (2017). Grading System and Student Effort. *Education Finance and Policy*, 12(1), 107-128.

UGC (2019). Evaluation Reforms in Higher Educational Institutions. *University Grants Commission.*

Johanson, George A. (1993). Absolute and Relative Grading Standards. *Annual Meeting of the American Educational Research Association, Atlanta, GA.*

Jon J. Denton (1989). Selecting an Appropriate Grading System. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 63:3, 107-110

George R. Quinn & Charles A. Szuberla (1963). Relative Grading. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 37(8), 490-494.

Nagaraju Dasari, Mahendra Parihar & Mohiddin Shaw Shaik (2021). Analysis of Exertion with Relative Grading Mechanism in Academics Using Computational Technique. *IOP Conf. Ser.: Mater. Sci. Eng. 1099* 012018.

G Beatrice Loka, Carmel McNaughtb and Kenneth Young (2015). Criterion-referenced and norm-referenced assessments: compatibility and complementarity. *Assessment & Evaluation in Higher Education, http://dx.doi.org/10.1080/02602938.2015.1022136,* Taylor & Francis.

Liz Freeman & Andy Miller (2001). Norm-referenced, Criterionreferenced, and Dynamic Assessment: What exactly is the point?. *The Educational Psychology in Practice: theory, research and practice in educational psychology,* 17(1), 3-16.