



Guru

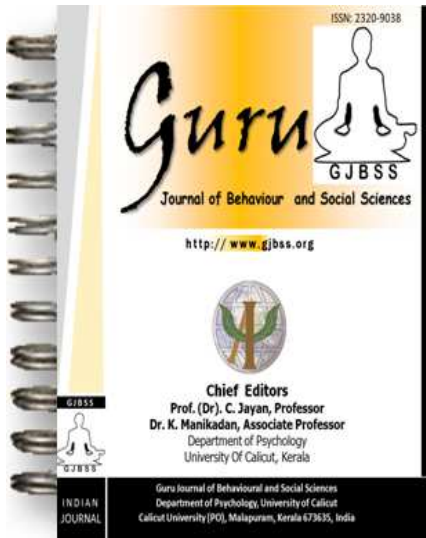
Journal of Behaviour and Social Sciences

Article Title

A Study of Reliability of Marking and Absolute Grading In Secondary Schools

Author(s)

Abdul Gafoor, K & Jisha, P



Chief Editors

Prof. (Dr). C. Jayan, Professor
Dr. K. Manikandan, Associate Professor,
Department of Psychology, University of Calicut



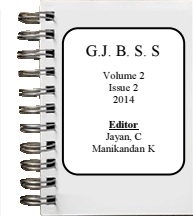
2014



Guru Journal of Behavioral and Social Sciences

Volume 2 Issue 2 (April – June, 2014)

ISSN: 2320-9038 www.gjbss.org



A Study of Reliability of Marking and Absolute Grading In Secondary Schools

Abdul Gafoor, K* & Jisha, P**

*Associate Professor, Department of Education, University of Calicut, Kerala, India

**HST (English), Government Girls HSS, Manjeri, Malappuram Dt, Kerala, India

Abstract

Received: 20 May 2014
 Revised: 25 May 2014
 Accepted: 2 June 2014

Keywords:

Assessment in Schools,
 Examination, Scoring,
 Consistency.

Using a non-experimental comparative group design in a sample consisting of 100 English teachers randomly selected from 30 secondary schools of a district of Kerala and assigning fifty teachers to groups for marking and grading, this study compares inter and intra- individual reliability in marking and absolute grading. Studying (1) the in marking and absolute Grading in terms of coefficient of rank correlation, (2) the intra-examiner and inter-examiner inconsistency in Marking and absolute Grading and (3) comparing Marking and absolute Grading in terms of intra and inter -examiner variation, revealed that grading system discourage teachers from looking for fine distinction where they do not exist and teachers honestly confess their inability to be precise in assessment while they score answer scripts intended for grading.

© 2014 Guru Journal of Behavioral and Social Sciences

Evaluation, a very important component of the education system, can fulfill or destroy purpose of education. Many policy documents pertaining to education from the Kothari Commission (1966) to the National Curriculum framework (2005) stated the inadequateness of the evaluation system especially the lack of full disclosure and transparency in grading and mark reporting. In any educational system, a major part of the evaluation of a student depends on the efficient conduct of examinations. The reporting of examination results is an important factor in the educational process. The symbols and signals employed in such reporting are usually marks, and the organization and operation of the process of establishing, recording, and transmitting such marks constitute the Grading system.

Preceding the recommendation of the National Policy on Education (NPE) 1986 on replacing marks with grades there was a nation wide debate. Introduction of grading, in fact, was as a reaction against the inadequacies of marks as a medium of evaluating certifying and reporting student competencies. Grades assume to give us an indication of the amount of learning and understanding and provide the feedback needed for the growth of knowledge and learning. One of the most important reasons for the failure of the examination system to achieve set purpose is the subjectivity on the part of the examiners. The problem of reliability in marking is not a new one. Among the factors which may be responsible for this unreliability in grading or marking are: (1) different standards of excellence both among different teachers and on the part of single teacher from one occasion to the next, (2) psychological factors, such as fatigue affecting ability to distinguish between closely allied degrees of merit, (3) systematic changes, and (4) the influence of handwriting. James and Sheppard (1971) for example have found positive relationship between the quality of handwriting and the grade accorded to test papers identical in content.

A mark is judgment of one person by another; it can provide both information and incentive" (Thorndike, 1977). As Ebel (1965) points out "there is nothing wrong in encouraging the student to work for higher marks, if, marks are valid measures of achievement. Marks provide motivation for better performance and encourage a student to bring out his best". However, marks have different meanings in different subjects and in different countries. Seventy percent may be a pass in USA where as in India 33 percent can be a pass mark. Marks



out of hundred in two different subjects are not equivalent and hence not comparable. Without transforming into standard scores, students and counselors for subject choice cannot use marks. Fifty percent in English language is not equal to fifty percent in Mathematics. 'Percentage marks have little meaning to those who have not seen the test used'.

Grades are for conveying students' performance with reference to specified criterion and the relative position of students in their group. Grading, a set of symbols or letters, categorize students based on their performance. Grading can be either direct or indirect. In indirect grading, marks are subsequently converted into letter grades using absolute standards. In this method of grading, it is possible that all students get grade A or grade E or any other grade in between. Norm referenced grading is based on relative achievement ie; comparison of achievement of a student with other students in the group (Grading on curve). These grades vary from year to year, school to school and subject to subject. Three, five, and nine point grading is used from lower to higher classes progressively. Communication of overall performance of the students is through Grade Point Average (GPA) in the grading system.

In the present day system, grading students' performance is more scientific as compared to marks. Several studies show that grading minimizes error in evaluation and is in vogue in most developed countries. Grades appreciate that precise measurement of human abilities is difficult, if not impossible. The major contribution of the grading system is to prevent fine distinction being looked for where they do not exist. Grading is an honest confession of our inability to be so precise in assessing human qualities. The system of Grading also has a few defects. Researches show that personality factors influence grades (Russell and Wellington, 1955; Hadley, 1954). Grading may put an eighty percent and an eighty-nine percent score in the same grade, say, grade two. Technically, they may be the same as their reference group but superficially as for an ordinary parent or future employer, eighty-nine percent would be better than an eighty percent.

It is the relativity of grades, which is their essential feature. When this feature is not understood properly, the method of "grading" is only "an illusory reform". A percentage mark is an "absolute", measurement of a particular students' performance on a particular test at a particular time. Therefore, marks "really depend more on test difficulty than on true quality of performance"(Lyman, 1986). A grade is an evaluation of mark, which is 'relative' to what psychologists and educational scientists consider to be a much more stable standard; the position of the student among his or her peers. This difference between "relative" and 'absolute' measurement is so very important to understand. After all, it would be hard to argue that merit in education is not a relative concept (NCERT, 2005). In this context, it is pertinent to investigate to what extent did; the introduction of grading solved the problem of reliability.

Rationale for publication after more than half a decade

Grades and marks communicate the achievement status of students to students, parents and the society at large. However, many arguments are placed against grading in schools. Literature shows that most of the criticisms against marking are equally applicable to grading. They include, the more people are rewarded for doing something, the more they tend to lose interest in whatever they had to do to get the reward (Kohn, 1993), "grade orientation" and a "learning orientation" are inversely related (Beck et al., 1991), that Grades tend to reduce students' preference for challenging tasks, encouraging cheating among students, and spoiling student- teacher and student-student relationships. In spite of such concerns, educational scenario in Kerala used to condense the discussion of this important aspect of formal education to a "marking vs. grading" match, and judge mostly in favour of the latter. This list surely cannot omit doubtful status of grading on validity, reliability, or objectivity too. It was in this context that this study was conducted in 2006-07. However the time was not ripe even for the community of educational researchers to accept a grading vs. marling comparison on equal terms.

Nonetheless, as years of grading in schools passes there is increasing force for the public outcry that the grades in schools are inflated and that this practice too does harm for the learning and development of younger generation. While there is no doubt that effective grading allows many students to actively update and advance their own learning, motivate them to attain the highest grades and to receive the recognition that accompanies such grades and to avoid the lowest grades, provide information to students for self-evaluation, for analysis of strengths and weaknesses, and for creating a general impression of academic promise, even the supporters of grading agree that teachers can teach without them and students can and do learn without them (Frisbie, & Waltman, 1992). Hence the authors feel that it is time to share findings of this study, to engender a discussion on further evolving the evaluation practices in schools for the better.

Key Terms

Marking is defined as a system which assigns a numerical score, used for evaluating and reporting achievement in students' work in schools, which was in vogue till the introduction of Grading System and even now being practiced as a preliminary step to Grading.

Absolute Grading is the system of grading where marks are subsequently converted into letter grades using absolute standards. In this study, the system followed by recognized secondary schools of Kerala (SCERT, 2005) is taken as example.

Table 1

Range of Marks, Grades and Grade Points as followed by Recognized Secondary Schools of Kerala

Range Of Marks	Grades	Grade Points
46-50	A ⁺	9
41-45	A	8
36-40	B ⁺	7
31-35	B	6
26-30	C ⁺	5
21-25	C	4
16-20	D ⁺	3
11-15	D	2
6-10	E	1

Reliability refers to the consistency or stability of the scores as evidenced through (1) the frequency of variation in two sets of scores,(2) the extent of variation in two sets of scores or (3) coefficient of correlation between a set of scores (Page, 1977).

Objectives

- To find out the inter-examiner consistency and intra-examiner in marking and absolute Grading in terms of coefficient of rank correlation
- To find out the intra-examiner and inter-examiner inconsistency in Marking and absolute Grading in terms of:
 - the percentage of teachers who are inconsistent during revaluation for marking /Grading; and
 - the extent of change in marks/ absolute grades during re-evaluation.
- To compare Marking and absolute Grading in terms of intra and inter -examiner variation with respect to:
 - the percentage of teachers who are inconsistent in Marking and Grading; and
 - the proportion of change in marks and absolute grades during re-evaluation.

Method

The method employed for comparing inter- and intra- individual reliability in marking and absolute grading was a non-experimental comparative group design. The details are as follows.

Participants

The participants for the study was 100 secondary school teachers teaching English drawn from 30 secondary schools of Malappuram district of Kerala using random sampling technique. Fifty teachers each were randomly assigned as sample for marking and grading.

Tools and techniques

1. Question paper in English language (standard X) of Half-Yearly Examination, in a government high school. It consisted of 31 questions including 8 objective type questions and 23 descriptive type questions. The maximum marks for the paper was 50. 2) Non-evaluated answer scripts of X Standard English Half Yearly Examination of the above high school. This served as the major tool for data collection. With the prior permission of the Head of the institution and the concerned English teacher, 35 answer scripts of a class were collected. Anything that discloses the identity of the examiner, name of the institution, name of the candidates, roll number, and other details were hidden and their Xerox copies (Four copies of each script) were taken. 3) Copies of marking scheme for the above question paper in required number .4) an open Ended Interview conducted by the investigator to collect the information on Difficulties felt in the area of evaluation and Suggestions for improving the present evaluation system from secondary school teacher.

Procedure

Collection of the data needed for the study was through the following steps.1) Teachers were randomly grouped into two, group for Marking (N=50) and group for Grading (N=50). 2) Answer scripts were allotted to the teachers such that in both groups (marking and grading) all the 50 teachers get the copy of same script twice (for estimating intra- individual consistency) and random 20 pairs of teachers in each group get the copy of the same script for evaluation (for estimating inter-individual consistency). 3) The teachers were asked to evaluate the answers with the help of the scoring key. Teachers belonging to marking group were told just to score the script following the answer key. Teachers belonging to absolute grading group were told that the scripts were meant for grading as being practiced in their schools , ie; they need to score as per the scheme and then turn the score into grades. Grading was done on a nine point scale with each letter (A+, A,B+,B,C+,C,D+,D and E). 4) All the teachers were informed that they would have to score one more script later on but none of them knew it would be the same scripts again. 5) The same teachers re-evaluated the same answer scripts after a period of two weeks and the results were collected the same way as earlier.6) An open-ended interview was also conducted by the investigator with each of these teachers in order to collect feed back regarding the present evaluation system, emphasizing the difficulties felt and suggestions for improvement.7) The scores obtained by re-valuation for both Marking and Grading were tabulated separately and later consolidated for the final analysis.

Results

Analysis of the data used Estimation of Percentage, Spearman's Rank Co-efficient of Correlation, and estimation of Fisher's 't' (Garrett,1979; Ferguson, 1976).

1. Consistency marking and absolute Grading in terms of coefficient of rank correlation

The correlation between the scores obtained for the same script while scored by the same examiner (intra examiner) and different examiners (inter examiner) were estimated in marking and grading groups . Results of estimation of coefficient of rank correlation are given in table 2.

Table 2

Results of estimation of coefficient of rank correlation showing intra and inter individual Consistency in marking and absolute Grading

	Marking		Absolute Grading	
	ρ	Fisher's t	ρ	Fisher's t
Intra-Examiner (N=50)	0.9986	130.75	0.9885	45.26
Inter-Examiner (N=20)	0.9974	55.76	0.9538	13.47

Table 2 shows that all the four coefficients of correlations are very high and highly significant. The co-efficient of correlation between the intra examiner marks ($r = 0.9986$; $t = 130.75$) and inter examiner marks ($r = 0.9974$; $t = 55.76$) are almost equal, with a high consistency in both. The extent of relationship is almost equal with a slightly higher correlation in the case of intra-examiner marking. Table 2 further shows that the coefficient of correlations between the intra examiner grades ($r = 0.9885$; $t = 45.26$) and inter-examiner grades ($r = 0.9538$; $t = 13.47$) are almost same with slightly lower correlation for inter-examiner Grading. The lowest of the correlations obtained is for inter-examiner grading, but the differences in correlations are not statistically significant.

2. Intra and Inter-examiner inconsistency in marking and absolute grading in terms of the percentage of inconsistent teachers and the extent of change in marks/ grades

Table 3 summarizes the percentage of teachers who are inconsistent while re mark or re grade, percentage of answer scripts for which the score/grade varied while evaluated by different teachers, and the result of comparison of marking and grading in terms of the relevant percentages (of teachers and scripts).

Table 3

Percentage of inconsistent teachers and the extent of change in marks/ grades during (intra and inter examiner) marking and grading

Group	Index Of Inconsistency Used	Intra-Examiner (N=50)	Inter-Examiner (N=20)
Marking	% of teachers/ answer scripts with change in score	52*	85**
Absolute Grading	% teachers/ answer scripts with change in score	24*	65**
Marking	Average deviation in score per script (%)	1.04	4.68
Absolute Grading	Average deviation in score per script (%)	3.55	11.11
Comparison Of Marking And Grading	in terms of percentage of inconsistent teachers (Critical Ratio)	9.71	1.46
	in Terms of Extent of change in Marks or Grades (Critical Ratio)	4.14	3.44

Note : *indicate percent of teachers showing intra examiner inconsistency

** indicate percent of answer scripts showing inter examiner inconsistency

Table 3 shows that there is a high inter-examiner inconsistency (85%) and intra-examiner inconsistency (52%) in Marking. The extent of average change in scores is also higher



in inter-examiner (4.68%) marking than in intra-examiner (1.04%) marking. As expectable, inconsistency is higher when two different examiners evaluate a script compared to the same examiner scoring it twice.

Data in table 3 reveals that in the case of grading also, intra-examiner variation is comparatively lesser than inter examiner-variation. Inter-examiner inconsistency in grading (65 %) is higher than intra-examiner inconsistency (24%) in Grading. The extent of average change in score during inter-examiner re-grading (11.11 %) is higher than the extent of average change in score during intra examiner re-grading (3.55 %). On the line of result obtained in marking, the number of teachers committing inconsistency during Grading and the extent of change in grades per answer script, both are higher in the case of inter-examiner grading compared to intra-examiner grading.

3. Comparison of Marking and absolute Grading in terms of intra and inter -examiner variation

Table 3 show that :

- (i) Difference in intra-examiner inconsistency between Marking (52%) and Grading (24%) in terms of with respect to the percentage of teachers who are inconsistent in revaluation is significant (CR =9.71) at 0.01 level.
- (ii) Difference in average intra examiner deviation between Marking (1.04) and Grading (3.55) during re-evaluation is significant (CR= 4.14) at 0.01 level.
- (iii) Difference in inter examiner inconsistency between Marking (85%) and Grading (65%) in terms of the percentage of teachers who are inconsistent in revaluation (CR = 1.46) is not significant at 0.05 level, and
- (iv) Difference in average inter examiner deviation between Marking (4.68) and Grading (11.11) during re-evaluation is significant (CR =3.44) at 0.01 level.

Conclusions

There is significant difference between Marking and Grading in both inter-examiner and intra-examiner inconsistency. The percentage of teachers committing inconsistency is higher in Marking while the extent of change in scores per answer script is higher in Grading. The highest consistency in scoring is in intra examiner revaluation for marking and the highest inconsistency is in inter- examiner revaluation for grading. The results suggest that grading system prevent fine distinction being looked for by teachers where they do not exist and teachers honestly confess their inability to be so precise in assessment while they score answer scripts intended for grading. However, this tendency to be less precise brings about huge variation (average variation = 11.11%) during inter examiner revaluation. This much variation per script surely affects not only scores, but the grades as well. This prompts the conclusion that teachers score answer scripts differently according to the purpose of scoring, ie; marking or grading. They try to be more precise and consistent while scoring scripts for marking than for grading.

Hence, the study has the following implications based on the suggestions made by teachers in the informal interview on difficulties felt and suggestions for improving the present evaluation system.

- 1) Administrators, decision makers or policy makers need to think about the implementation strategies of Grading rather than stressing the theoretic merits of Grading over marking. 'Grades are not the panacea that have made them out to be, and the transition from marks to grades is a minor exam reform at best' (NCERT, 2005). A shift from absolute to relative grading may be desirable.
- 2) Concrete directions regarding the means to make grading more objective and scientific need to be thought upon and implemented.
- 3) In spite of large-scale in-service programmes conducted to acquaint the grading practices, teachers continue to complaint that the basic concept of Grading is not comprehensible to them. Steps to make the teachers aware of the theoretical and practical aspects of Grading need continuation.
- 4) Time



for evaluation should be considered and given adequate weightage in calculating the workload of teachers. 5) Adequate time for evaluation and concrete and scientific methods of solving the issue of borderline cases while grading must be provided. While grading extra-vigil is to be taken in evaluating border-line cases. 6) Teachers must be made aware that subjectivity equal to or more than marking will affect Grading and so they need to be alert through out the evaluation procedure and 7) The sanctity of grades should not be over emphasized since grades are derived from marks. 8). In-service and pre-service Teacher education programmes have to demonstrate the uses, the strengths and the weaknesses of different grading practices like Relative grading, like Grading on the Curve, Distribution Gap Method, Standard Deviation Method or absolute grading like Fixed Percent Scale, Total Point Method, Content-Based Method, (Frisbie, & Waltman, 1992) by going beyond the age-old argument that grading is better than marking. 9). Teachers need be made aware that employing descriptive rubric makes grading more informative, reliable and valuable. 10). Evolving better grading practices require that educators be allowed to experiment with grading practices even as they are allowed flexibility in choosing the educational objectives, processes, experiences and resources; though surely not at the grades in which public certification is required.

References

- Beck, H. P., Rorrer-Woody, S., & Pierce, L. G. (1991). The relations of learning and grade orientations to academic performance. *Teaching of Psychology*, 18(1), 35-37.
- Ebel, R. I., & Frishie, D. A. (1991). *Essentials of Educational Measurement* Eaglewoods Cliffs.
- Ferguson, G. A. (1976). *Statistical Analysis in Psychology and Education*. Tokyo : Mc Graw Hill, Kogakush.
- Frisbie, D. A., & Waltman, K. K. (1992). Developing a personal grading plan. *Educational Measurement: Issues and Practice*, 11(3), 35-42.
- Garrett, H. E. (1979). *Statistical techniques in Psychology and Education*, Bombay : Bakils, Fetter and Science (Pvt.) Ltd.
- GOI (1966). *Report of the education commission: education and national developmen*. New Delhi: Ministry of Education.
- Kohn, A. (1993). Why incentive plans cannot work. *Harvard business review*, 71(5), 54-60..
- Lyman, H. B. (1986). *Test scores and what they mean*. Englewood Cliffs, NJ: prentice hall
- NCERT (1985). *Contemporary Issues in Public Examination*, Publication Department, Aurobindo Marg, New Delhi.
- NCERT. (2005). *National curriculum framework 2005: Position paper on Examination reforms*. New Delhi: Publication Department, NCERT.
- Page, et al. (1977). *International Dictionary of Education*. New York : Harper.
- SCERT. (2004). *Continuous Evaluation: Content and Application*. Trivandrum : DPI.
- SCERT. (2005). *Padavugal: A Handbook for Evaluation*, Trivandrum : DPI.
- Thorndike ,R., & Hagen, E. (1977). *Measurement and evaluation in psychology and education*. New York: Macmillan