

SCALING IN EXTERNAL EXAMINATIONS

Introduction

The scaling or standardization of marks in external examinations has been an issue of much concern to many students, parents, and members of the public at large. Many such concerns arise out of misconceptions of what scaling is, how it is done, and the rationale for its use. One only has to read the daily newspapers immediately after examination results are released each year to know of the many expressions of dissatisfaction by students and others and allegations of unjust practices due to the scaling of marks.

When concerns such as these are not adequately addressed or are left unanswered they tend to create doubts which can lead to suspicions and a loss of confidence in the examination system. Such a situation must be avoided as it has led many to exploit the sensitivity and fragility of examinations for their own ends.

Because many such allegations are channeled through dissatisfaction over marks, one wonders whether the scaling process has been, in such cases, used as an excuse for a student's failure to achieve what he or she expects in an examination. Whatever the case may be, there is obviously a need to make students and the public at large, aware of the need for scaling in external examinations and how it is done. An examination is, after all, a high stakes event in the life of a student.

Common Questions

A number of questions are commonly raised with regard to scaling. These include:

1. Are marks scaled? If so, why?
2. Does scaling reduce marks?
3. Why are marks given in internal school examinations not consistent with external examination marks?
4. Is there discrimination in scaling?
5. How is scaling done?

The following discussion seeks to address the above questions and clarify other relevant issues which are critical to a basic understanding of this essential process in external examinations.

Are marks scaled?

The process of scaling has always been a part of any external examination that is used to rank students for selection. It was introduced locally as the Ministry began to set and process its own external examinations. The New Zealand examinations that our students sat years ago were standardized and so were other external exams elsewhere that were used for ranking students.

Our neighboring Pacific islands like Tonga, Samoa, Vanuatu, Solomon Islands, Nauru, Kiribati, and Tuvalu have their external exams standardized for exactly the same reasons that we have. In fact, our current scaling system is the same as that used in these countries as well as New Zealand. Our experiences in Fiji are not unlike theirs with regard to scaling.

Similar questions are asked and suspicions expressed although the islands may not exploit the media in this regard as much as we do here.

Even if students were not to be ranked, a certain amount of scaling would still need to occur to address the issue of comparability between subjects.

Subject Equivalence and Comparability

Without comparability, it is difficult to achieve fairness in an external examination. This comparability is based on the principle of subject equivalence, an idea which is embraced by most responsible educational authorities. The latter states that all subjects offered at one level of schooling are assumed to be equivalent for assessment purposes (Johnston & McCaw, 1996). What this means is that it is as easy or difficult for a student to score a mark of, say 60%, in one subject as in another.

The principle of subject equivalence or comparability is easily reflected in an examination when all students are sitting all the subjects. How is this so? When all students are sitting all the subjects, the range of abilities in all the subject cohorts can be said to be similar as the same group of students is involved. In other words, the norm in the groups is similar. This is also an essential feature or condition for ensuring the validity of the measurement in question, an external examination being a norm-referenced type of assessment. While some adjustment of scores may still be needed in such a case, it is unlikely to change the original scenario or the raw distribution of marks to a great extent. In other words, marks may change, but only slightly, through scaling when all students are sitting all the subjects.

Scaling by the Overall Mean and Overall Standard Deviation

The above situation is hardly the case in Fiji. Our students are given choices of subjects to sit apart from one or more compulsory subjects. This leads to a situation that is unlike that which is described above. In this case, the norm in the various subjects is different and to achieve subject equivalence or comparability is not easy. To address the situation, scaling assumes that all students take all subjects. Based on an overall mean and overall standard deviation (values mandated by the examining authority) for the whole examination, adjustments are made to all subject distributions to bring their mean and standard deviations to the overall/national values. In performing their task in the process, the overall mean and overall standard deviation serve to maintain the overall standard of the examination. These values therefore remain the same from year to year.

As this adjustment is made, marks may increase, decrease or not change at all.

Other Scores and Relative Standing

In the same process of determining marks, two factors are also taken into account, These are a student's scores in his other subjects as well as his relative standing in the subject cohort. These two factors together with the overall mean and standard deviation are three parameters determining the final outcomes of standardization.

Level of Difficulty of Papers

The inconsistency described above that is due to the varying number or abilities of students in various subjects is only one example of the many sources of variation in an examination. Another significant source of variation is the differences in the difficulty level of the various papers. While the intention in an examination is to set all papers at the same level of difficulty, the raw examination results usually tell otherwise, as some papers are found to be easier than others. This is shown by the varying means and standard deviation values of the raw distributions of scores. This inconsistency can be addressed by the scaling process described above.

Scaling also serves to restore the comparability that is lost through other sources of variation in an examination.

Combining Scores

Another reason for scaling is to enable subject scores to be meaningfully or validly combined to produce an aggregate score (as in English + Best 3 for FSLC). The aggregate score is then used as an indicator of a student's overall achievement. As such, the aggregate score is used to rank students for scholarship selection or entry into tertiary institutions.

For scores to be combined or aggregated, they must first be made to be "similar" (Pongi V. 2000). Just as X Australian dollars cannot be combined with X Fiji dollars unless the two amounts are of the same currency, a meaningful aggregate cannot be achieved unless the scores are first made to be "similar". To achieve such similarity in combining scores, the distribution of scores in the various subjects must be adjusted, usually to the same means and standard deviations. Only then can subject scores be validly aggregated. The same process would enable the subject scores to be equally weighted in the aggregate. The equal weighting of subjects in an aggregate is necessary for an objective assessment of a student's overall achievement. This process is critical in achieving a valid ranking of students for selection and in enabling the most deserving students to be awarded scholarships.

Purposes of Scaling

Standardization, therefore, occurs for three basic reasons:

1. To remove the negative influences of the many sources of variation in an examination and thus bring about comparability.
2. To standardize the examination from year to year.
3. To enable a valid aggregation of scores which are used to rank students for selection.

Non-discriminatory

Scaling is applied equally to all students at the same time and in conditions of anonymity. It is therefore not discriminatory in any way at all. The claim that certain ethnic groups

have been adversely affected by standardization is baseless and should never be entertained.

Comparability with Internal Examination Marks

The marks gained by a student in an internal examination cannot be compared with those obtained in an external one. This is because the norm in the two examinations are not the same. While the marks for a subject may be the same or close in the two examinations, it does not follow that the marks are comparable.

Conclusion

Scaling is, basically, a principled way of adjusting the distribution of scores in an examination for the ultimate purpose of achieving validity and fairness. These qualities determine how good a test is and they are not easily achieved given the many sources of variation in the examining process. Such variations can be due, for instance, to differences in the level of difficulty of papers, inconsistencies in marking and differences in the number of students appearing for each subject.

It should be noted that, by its very own nature, an examination exerts a negative influence on itself through the many human factors at play in its development. Because human factors are not easily controlled, they become potential sources of variation. By adjusting the distribution of scores in the various subjects examined and making them similar on the basis of their means and standard deviation, scaling helps to remove the negative effects of such inconsistencies and restore the standards required for each subject and that for the overall examination.

The process of scaling is also necessary to maintain the standard of an examination as a whole from year to year. This is made possible by the use of an overall mean and standard deviation.

Because most external examinations are used for the purpose of selection, scaling in these examinations is an essential process to ensure the valid aggregation of scores and in turn a valid ranking of students for selection purposes.

References:

Johnson & McCaw, 1996. "Review of Fiji's Examination System".

Pongi V., 2000. "Combining Assessment Scores; Is it a necessary Procedure?" (Unpublished)
- Paper Presented at the Fiji Principals' Association Conference 2000.

By Salote Rabuka
Examinations Office, Ministry of Education