

Linking school-based assessment with public examinations: Role of moderation – principles, policy and practice

Gabrielle Matters

Summary

With the increased emphasis on teacher judgment in high-stakes assessment has taken the role of moderation in ensuring that student performances of equivalent standard are recognised as such. The following situations demand that moderation occur.

- Before subject results can be recorded on a certificate with comparability guaranteed; and,
- In ensuring that the results of different assessments (different subjects, different sites) are on a common scale before being combined to produce a tertiary entrance rank.

The full paper outlines the theoretical underpinnings of moderation, and describes applications of these in terms of form, purpose, and technique. The focus of the paper is on the use of statistical moderation to scale school-based assessments to public examination results.

One of the defining characteristics of a jurisdiction's assessment arrangements is the nature of the regime that sets and marks the components of an assessment program for which student results appear on a certificate. There are many possible variations. In what follows, these variations are presented as a series of questions and answers.

1. *Who sets the assessments? Who composes the marking schemes? Who marks the student work?*

Here are the possibilities.

	Who sets the assessments?	Who composes the marking schemes?	Who marks the student work?
Public examinations	E	E	E
School-based assessments	I	I	I
Hybrid model	E	E	I

Legend:

E = external source (agency such as examining board)

I = internal source (teachers/schools)

2. *In the case of pure school-based assessment (I, I, I) and of the hybrid model (E, E, I), what mechanism is used to respond to the reliability challenge arising from an internal locus of control?*

The possible forms of validating teacher judgments at the marking stage are statistical moderation and social moderation. Linn (1993) refers to moderation involving human judgments as social moderation. He contrasts this with statistical moderation where adjustments are made to sets of scores to make them comparable or equivalent in some way. An example of statistical moderation is adjusting a school's internal assessment scores in a subject against the distribution of external examination scores for that group of students in that subject. Statistical adjustment of this kind is best referred to as "scaling".

And there are variations within each of these techniques in practice.

3. *In the case of social moderation, which technique is used?*
 4. *In the case of statistical moderation, which technique is used?*

The available techniques for each form of moderation are tabulated below. (Entries in the table are explained in detail in the full paper.)

Purpose	Form	Technique	
Validation of teacher judgments in school-based assessment	Social moderation (Consensus within the "guild of professionals")	Teacher meetings (peer review)	
		Review panels (expert review)	
		Visitation (moderators) (bureaucratic review)	
		Statistical moderation (Scaling)	External examinations
		Aptitude test	
	Omnibus test		
	Other subjects (iterative)		

5. *What formula is applied in scaling?*

Statistical moderation is the process of adjusting the school-based assessment distributions for a subject so that the average and spread of the school-based assessments match the average and spread of the school's distribution on the external examination scores for the subject.

In technical terms, it uses a linear transformation to adjust the school-based assessment distribution for a subject in a school to have the same mean and standard deviation as the distribution of examination results for that subject achieved by students at that school.

The linear transformation does not change the rank order of students or the relative differences between them. For a given student, adjusted scores thus obtained (called “scaled score”) may be higher or lower than the original school scores depending on the location and spread of the school scores.

The following equation is used to scale school-based assessments.

$$MSA_{\text{student 1}} = \left\{ \left[\frac{(SBA_{\text{student 1}} - \text{Mean } SBA_{\text{school}})}{SD \text{ } SBA_{\text{school}}} \right] \times SD \text{ } E_{\text{school}} \right\} + \text{Mean } E_{\text{school}}$$

Where $MSA_{\text{student 1}}$ is the moderated school-based assessment (scaled score) for student 1 in the school for a subject;

$SBA_{\text{student 1}}$ is the school-based assessment for student 1 in the school for a subject;

$\text{Mean } SBA_{\text{school}}$ is the mean or average of the school-based assessments for the school in the subject;

$SD \text{ } SBA_{\text{school}}$ is the standard deviation of the school-based assessments for the school in the subject;

$SD \text{ } E_{\text{school}}$ is the standard deviation of the examination marks for the school in the subject; and,

$\text{Mean } E_{\text{school}}$ is the mean or average of the examination marks for the school in the subject.

Limitations and advantages of statistical moderation

Since the two sets of marks refer to different assessment performances, the performances may not reference the same underlying characteristics. Furthermore, even if they did reference the same characteristics (e.g. where the school assessments mirror the external assessments in their content and form), the standard of the performances could be quite different. Therefore, the scaling merely realizes an expectation that one set of results mirror the other set of results. The process of scaling adopts the assumption of equivalence but cannot verify it.

Statistical moderation is the cheapest method for achieving comparability. It does not involve movement of materials (samples of student work) or movement of personnel (teachers, panelists, moderators). It is efficient because it relies on algorithms and computer processing of data.

Other limitations and advantages are discussed in the full paper.

References

This paper draws on the writings of others who are cited in the full paper.