# PRINCIPLES OF MODERATION OF INTERNAL ASSESSMENT

Anthony Daly, Lucy Billington, Suzanne Chamberlain, Lesley Meyer, Neil Stringer, Martin Taylor, Kate Tremain

## INTRODUCTION

Assessing students' work and awarding a mark requires human judgement. For example, two people assessing the same piece of work may disagree on whether a skill has been demonstrated to a particular level or whether a student's written answer is clear. Or, a person may mark and then re-mark a piece of work at a different time and make slightly different judgements. While it is possible to reduce differences between and within individual assessors with practice, experience and training, processes need to be in place to ensure that assessment decisions are valid, reliable, fair and consistent with national standards. For this reason the moderation of internal assessments takes place to ensure that assessment decisions are in line with required standards.

This chapter will outline a variety of types of moderation and verification of internally-assessed work. It will discuss issues relating to, for example, accuracy and validity, quality assurance and control, and cost effectiveness. It provides some operational comparisons between different forms of moderation and verification currently in use, while considering some of the advantages and disadvantages of different types of moderation.

## QUALITY OF ASSESSMENT

Quality procedures are used to improve the inherently imperfect process of assessment. Although these quality procedures can differ slightly in the way that they occur within moderation activities, all are focussed upon the setting, taking and marking of assessment tasks. The box below outlines the minor differences in how quality assurance and quality control are defined, as they relate to moderation.

| Quality assurance | Quality control |
|---|---|
| ▪ a feed-forward (pre-outcome) approach with a focus on prevention | ▪ a feed-back (post-outcome) approach with a focus on detection |
| ▪ monitoring steps to avoid substandard assessment activities and products | ▪ scrutiny of assessment outcomes to ensure fairness and consistency in assessment procedures |
| ▪ processes taking place within a school through internal standardisation, moderation, or verification | ▪ external to the school through awarding bodies' moderation or verification procedures |
| ▪ determination of whether proper procedures are in place to provide accurate assessment | ▪ determination of whether outcomes are satisfactory |
| ▪ relates to assessment criteria and their meanings, rather than the assessment itself | |
| ▪ processes ensure assessments are fit for purpose | |

There isn't always a clear distinction or separation between quality assurance and quality control. For example, there can be interactions between the different methods of evaluating the quality of assessment processes and products, so that the results of quality control procedures can help inform quality assurance

procedures. Broadly speaking, then, moderation can be considered a form of quality *management*, something which provides accountability and a means of improving assessment.

## TYPES OF MODERATION

In reality, teachers are unlikely to apply identical standards and criteria when marking school-based components of examinations, so moderation procedures are required to check and, if necessary, adjust teachers' marks. There are two main types of moderation: *statistical* moderation and *social* moderation. Statistical moderation aims to match up performances from different sources or in different subject areas by statistically adjusting scores in an attempt to make them comparable. For example, between-teacher (or school) comparisons can use an external examination as an anchor measure to statistically moderate scores assigned by teachers to the students at each school, such that results may change without altering the relative standing of individual students within a school.

Rather than using statistical techniques to align standards, social moderation relies upon professional judgement and consensus as to what constitutes a performance which would meet those standards. Hence, social moderation relies primarily on judgement, although statistical methods are sometimes incorporated on a secondary basis, as is discussed below. Types of social moderation include, for example, approaches where individual performances on distinct tasks can be rated and compared using a common framework and standard. Alternatively, scores assigned by different teachers within a school, or a cluster of schools, on the same tasks can be compared with a national standard and adjusted accordingly. Importantly, and in contrast to the statistical approaches, social moderation has a greater emphasis on collegiality and the achieving of consensus through social interaction and staff development. The next sections address statistical and social approaches to moderation in greater detail.

## Statistical Moderation

As stated above, statistical moderation aims to compare scores by statistically adjusting the scores; it takes place after the assessment process. An external test may be used as an anchor to statistically moderate scores between schools, because the external score is based on a test that is taken by all students across the country, region or, in our case, awarding body: it provides a common standard against which the schools-based assessments can then be compared. This process adjusts the level and spread of each school's assessments of its students' performances, to match the level and spread of the same students' scores on the common external examination.

Table 1 below illustrates a relatively simple example of statistical moderation and shows the results of 12 students, six from each of two schools. Let's assume students in both schools were taking the same subject, but were assessed using tasks that were internally set and marked. In other words, these tasks could differ between schools yet still be broadly equivalent in terms of the curriculum. The students were given scores by the teacher and then rank ordered according to how they performed in relation to their classmates. Students from both schools were then given a common assessment task that was externally set and marked and taken under standardised conditions. This assessment task could have been a general achievement test (e.g., SATs in the UK) or a common, externally-set unit within the subject. Students' scores on the external assessment were then combined with their internal scores to produce a combined score. A grade was then awarded from this score, but it was based on the common component. This process allows all students to receive a grade that compensates for possible variations between schools in the internal assessment. In this example, students' rank order within the school is not carried across to their final overall rank, although other methods of statistical moderation might maintain the rank ordering of students within a school. In school B, student T was ranked fourth in the school after the internal assessment. However, T performed very well in the external assessment which compensated for the relatively much poorer

2

performance in the internal task. As highlighted in the table below, this resulted in student T receiving a higher final rank (second in school B and third overall), with students R and S each effectively dropping down the school order by one position relative to their rankings after the internal assessment.

**Table 1. An example of statistical moderation.**

| School | Candidate | Internal assessment | Rank order | External assessment | Combined score | Overall rank | Grade |
|--------|-----------|---------------------|------------|---------------------|----------------|--------------|-------|
| A | K | 80 | 1 | 130 | 210 | 1 | A |
|   | L | 70 | 2 | 90 | 160 | 4 | B |
|   | M | 60 | 3 | 80 | 140 | 7 | C |
|   | N | 50 | 4 | 85 | 135 | 8 | D |
|   | O | 40 | 5 | 75 | 115 | 9 | E |
|   | P | 35 | 6 | 60 | 95 | 11 | E |
| B | Q | 85 | 1 | 115 | 200 | 2 | A |
|   | R | 75 | 2 | 80 | 155 | 5 | C |
|   | S | 60 | 3 | 90 | 150 | 6 | C |
|   | T | 55 | 4 | 115 | 170 | 3 | B |
|   | U | 45 | 5 | 65 | 110 | 10 | E |
|   | V | 40 | 6 | 50 | 90 | 12 | E |

The types of anchor instruments, also referred to as reference tests or external calibrators, may include tests of general achievement, aptitude, or intelligence. However, there are risks associated with using a reference test to statistically manipulate teacher ratings. If the reference test itself has low reliability or is biased towards certain groups, then the adjusted results will also have low reliability and be biased. In addition, a teacher assessment and a reference test may not share the same assessment objectives, although there should obviously be some overlap to provide a reasonable basis for moderation. Finally, although statistical moderation may determine the final assessment outcomes, the process deals directly with scores and is not dependent on the evidence from which those scores were actually derived.

Statistical moderation can be used to *calibrate* or to *monitor*. Calibrating refers to the process, as described above, whereby statistical adjustments are universally applied to teacher assessments based on a reference test. Monitoring refers to a broader approach in the use of statistical comparisons, where moderation is used to detect comparability issues by, for example, triggering an inspection. It may also be used for longer term monitoring of schools or awarding bodies. There are very few instances of calibration, whereby statistical moderation is the sole method used to adjust the scores from an internal assessment.

Statistical moderation is not without its difficulties. For example, AQA conducted a study to model the effects of a range of statistical moderation techniques on outcomes of a number of GCE and GCSE examinations. Four methods of statistical moderation were modelled:

1. Adjusting the internally-assessed marks to have the same mean as the moderating instrument.
2. Adjusting the internally-assessed marks to have the same mean and standard deviation as the moderating instrument.

3. Regressing the moderating instrument marks onto the internally-assessed marks, using the regression line to adjust the internally-assessed marks.
4. Mapping students' moderated internally-assessed marks to the moderating instrument marks, keeping the same rank order.
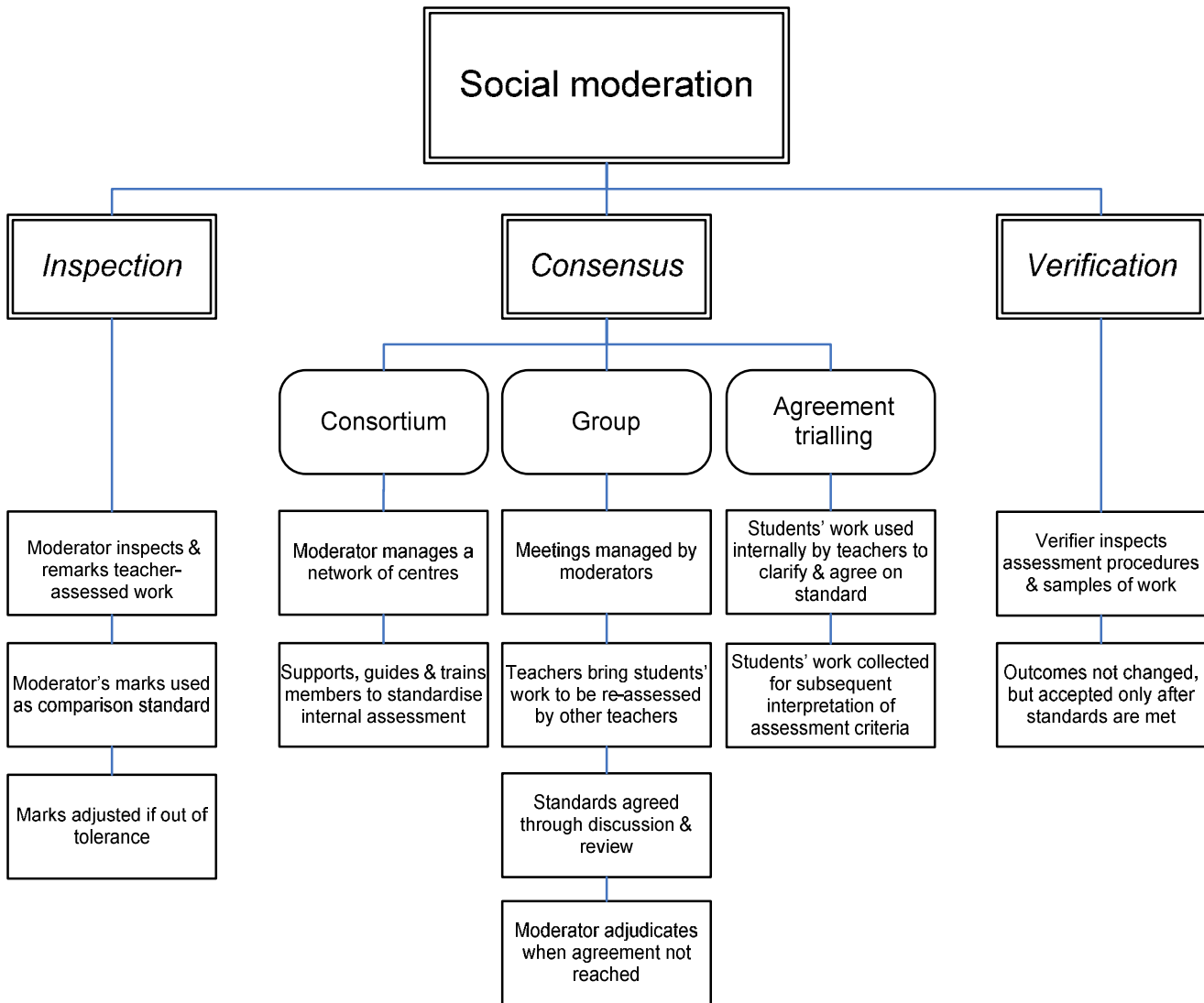
There were no great differences found between the four methods of statistical moderation. However, the outcomes using these statistical methods varied considerably from the 'real' outcomes that had been awarded using the existing system of moderation. In addition, the outcomes of all four methods of statistical moderation did not produce consistent results across different subjects. These findings suggest that statistical moderation is not a viable alternative in the current GCE and GCSE system.

Statistical moderation is costly in terms of designing and maintaining effective and appropriate external tests, although this is less of an issue where an external test pre-exists as a component of an examination. This type of moderation is generally poorly understood by teachers (and almost everyone else). In addition, given that statistical moderation is rather removed from the assessment process itself, there is no feedback available that teachers might use to inform and modify their assessment practices. Teachers, and others, often question the appropriateness of using written papers as a moderating tool for practical work or for performance of an ephemeral nature. Finally, in countries where statistical moderation is employed, it is usually not the sole form of moderation as it is used in combination with some form of social moderation and as a means of monitoring.

## Social Moderation

There are three main types of moderation that are based on a social rather than statistical approach. As Figure 1 over the page shows, the three types are: moderation by *inspection*; *consensus* moderation, which includes group moderation, agreement trialling and consortium moderation; and *verification*.

*Moderation by inspection* is based on moderators inspecting the teacher-assessed work of a sample of students from each school. This work is typically posted or, increasingly, sent electronically to moderators. Student work that doesn't lend itself to these methods, such as artefacts or performances, may require the moderator to visit the school. Sampling procedures, which are generally determined by the awarding body and moderator, aim to include work which covers the full range of expected marks. The work is then re-marked by the moderator and these "moderator marks" are then used as a comparison with the teacher's marks. If a school's marks differ from the moderator's marks, falling outside predetermined tolerance limits, adjustments are made to the school's marks to bring them in line with the moderator marks. Generally, when moderator adjustments are made, the rank ordering of students within each school is maintained. While there may sometimes be a statistical component associated with this procedure within some awarding bodies, it will merely be the automating of what would normally be a judgemental process of bringing school and moderator marks in line. For example, although linear regression could be used as a tool to align school and moderator marks, the process still relies on the re-marking of a sample of work by a moderator and it is, therefore, not strictly statistical moderation.

```
                          ┌─────────────────────────┐
                          │                         │
                          │    Social moderation    │
                          │                         │
                          └─────────────────────────┘
                                     │
        ┌────────────────────────────┼────────────────────────────┐
┌───────────────┐          ┌───────────────┐          ┌───────────────┐
│   Inspection  │          │   Consensus   │          │  Verification │
└───────────────┘          └───────────────┘          └───────────────┘
```

Figure contents:

**Inspection**
- Moderator inspects & remarks teacher-assessed work
- Moderator's marks used as comparison standard
- Marks adjusted if out of tolerance

**Consensus**

*Consortium*
- Moderator manages a network of centres
- Supports, guides & trains members to standardise internal assessment

*Group*
- Meetings managed by moderators
- Teachers bring students' work to be re-assessed by other teachers
- Standards agreed through discussion & review
- Moderator adjudicates when agreement not reached

*Agreement trialling*
- Students' work used internally by teachers to clarify & agree on standard
- Students' work collected for subsequent interpretation of assessment criteria

**Verification**
- Verifier inspects assessment procedures & samples of work
- Outcomes not changed, but accepted only after standards are met

**Figure 1.  The three main types of social moderation.**

*Group moderation* requires teachers to bring samples of their students' work to a meeting in which the work is reassessed by other teachers. These meetings are usually conducted by awarding bodies or their appointed moderators who guide and manage the review process. The aim is to ensure that judgements are comparable by using a process of discussion and review, so that there are similar interpretations of standards and that there is a similar recognition of the performances that represent those standards. This is an active process, with an aim to reach a consensus and to reconcile any disagreements. Where agreement cannot be reached, usually the moderator will adjudicate and decide at which point a sufficient degree of agreement exists.

*Agreement trialling* and *consortium moderation* are variations on the theme of consensus moderation. Agreement trialling is a process which places greater emphasis on the professional development of assessors. Students' work is used as a basis for clarifying assessment criteria, with assessors identifying the critical points in the process of deciding what represents the standard which, in turn, gives them the chance to agree upon that standard. For example, teachers in a school department would collect and analyse students' work in a portfolio to determine the standards, storing examples to provide a reference point for subsequent interpretations of the assessment criteria.

Consortium moderation refers to systems in which, for example, a network of schools is managed by an adviser (moderator) to provide support and guidance to teachers in the delivery and assessment of a specification. The adviser assists in the training of consortium members in the assessment of coursework and provides guidance on ensuring effective internal standardisation of marking within the schools. The *area moderation* process in AQA's GCSE Art and Design qualification provides an example of moderation that combines elements of group and consortium moderation. In the AQA area moderation system, the standards are controlled by the moderator, rather than being determined through the consensus of the teacher-assessors.

*Verification* covers a variety of processes and is normally used in vocational qualifications, where the assessment process determines whether a student has met a set of criteria, rather than *how well* those criteria are met. At one end of the spectrum, it is very similar to moderation by inspection, with the main difference being that the external verifier cannot normally change the school's assessment decisions. A verifier will attend a school and check that it is operating according to established standards by inspecting the school's assessment procedures and samples of student work. If the verifier decides that standards are not being met, appropriate feedback and guidance are given to the school, possibly through a school's quality manager. Although the school's assessment decisions are not changed by the verifier, the school's decisions may not be ratified until the required standard is reached. At the other end of the spectrum, verification can refer to a simple checking process to verify that correct procedures have been followed without a judgement of whether assessment standards are met.

| Advantages | Disadvantages |
|---|---|
| *Inspection* | |
| ▪ emphasis on external control | ▪ focus on the outcome rather than the process |
| ▪ greater consistency in the interpretation and application of standards | ▪ possible reduced emphasis on validity |
| ▪ reliability of teacher assessment maximised | ▪ little opportunity for professional development |
| ▪ possible to moderate written coursework by post or electronically | ▪ assessment innovation potentially limited by external control |
| | ▪ moderator visits to schools typically required for artefacts and performances |
| *Consensus* | |
| ▪ greater emphasis on validity | ▪ can be very costly |
| ▪ greater control for teachers | ▪ greater burden on teachers |
| ▪ more professional development opportunities for teachers | |
| ▪ innovation in assessment fostered | |
| ▪ reduced emphasis on external testing | |
| *Verification* | |
| ▪ straightforward | ▪ focus on process rather than outcomes |
| ▪ low cost | ▪ lack of common "standard" |
| ▪ reduced demand on schools | ▪ assessment outcomes not comparable |

In terms of reliable assessment outcomes, there are clear advantages to moderation by inspection, due in part to its emphasis on external control. There is greater consistency in the interpretation and application of standards and a maximising of the reliability of teacher assessment. Disadvantages include problems associated with a focus on the outcome rather than the process and a decreased emphasis on validity of assessment practice generally. There can be little opportunity and support for professional development of assessors and a risk that external control can limit innovation in assessment. The nature of the assessment itself has an additional impact. For example, whereas written coursework can be moderated by post, artefacts and especially student performances are more problematic, with substantial costs involved in ensuring moderators are able to visit and adequately cover all schools. These costs include travel, accommodation and logistical problems associated with accessing all schools within a large geographic area.

Consensus moderation is widely considered to be the most effective way of promoting professional development for teachers and fostering innovation while achieving the required levels of assessment standards. However, it is very costly in terms of finances and resources, with greater burden placed on teachers, although this burden may be offset to some extent by a reduced emphasis on external testing and the benefits of professional development. The verification process suffers from criticism that it does not allow comparability of assessment outcomes between students or from school to school. This has resulted in some verification procedures placing a greater emphasis on those aspects more common to moderation by inspection. The verifier role has also become more complex with, for example, a greater emphasis on training and subject specialisation.

## EXAMPLES OF MODERATION

### Statistical Moderation

Currently, no UK awarding bodies use an external assessment component to statistically moderate internally-assessed marks. Therefore, we'll look briefly at what happens in Australia, because the most sophisticated and elaborate methods of moderation can be found there, leading to Australia being considered the home of statistical moderation methods. The Australian education system is run on a state-by-state basis and there are currently nine separate certificates of secondary education. Below is brief outline of how variations of the statistical and social moderation are used in combination in Australia.

---

**Australian Capital Territory:** consensus and statistical moderation, with the latter used for determining university entrance scores.

**New South Wales:** statistical moderation and within-school moderation, with external examination (50% of overall assessment) used to align internal assessment.

**Queensland:** entirely consensus moderation, with a core skills test of cross-curriculum skills used for tertiary entrance. This test is taken only by those students who intend to apply for a university place.

**South Australia:** uses group moderation with subject experts and school-based moderation (with moderator visits for student performances). Statistical moderation is used in subjects that have an external assessment, with these marks used to align assessments moderated at the school level.

---

> **Tasmania:** consensus moderation with statistical monitoring.
>
> **Victoria:** external examinations and statistical moderation to align internal assessment. There is also a General Achievement Test which is used as "predictive" quality assurance and monitoring.
>
> **Western Australia:** consensus moderation with statistical moderation for monitoring. The external assessment (50% of overall assessment) is taken only by those students who intend to apply for a university place; those not seeking a place are assessed only through their internally-assessed marks.

## Social Moderation

All UK awarding bodies employ some form of moderation by inspection (for general qualifications) or verification (for vocational qualifications), using a variety of methods to monitor and/or moderate marks from internal assessment. Figure 2 below shows the typical structure of a GCE or GCSE moderating team, beginning with the teacher as assessor on the right and ending with the subject's Chair of Examiners who, through the Chief Examiner, may have responsibility of a number of Principal Moderators and Principal Examiners. Assessment by teachers will take place under some form of internal standardisation within each school in line with awarding body guidance, with a team of moderators each having responsibility for a number of schools. Coursework advisers (who are also moderators) lead standardisation meetings for teachers, provide advice to schools and give feedback to the Principal Moderator.
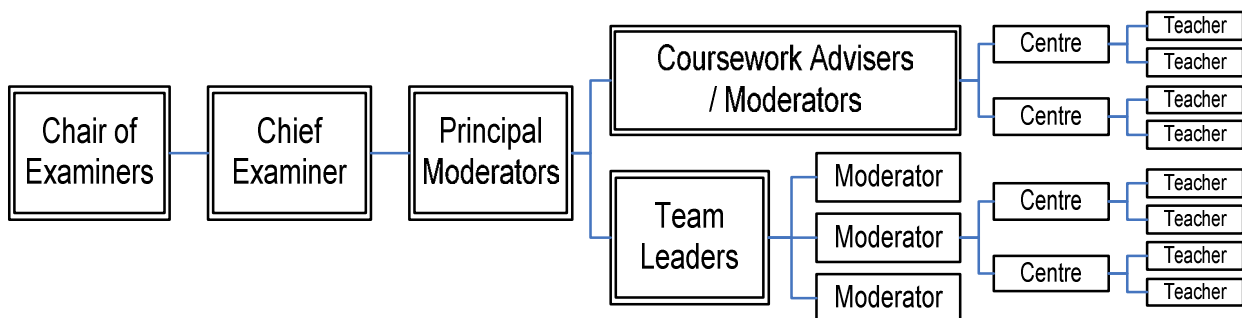


**Figure 2: Typical structure of a moderation team in a general qualification (GCE or GCSE).**

Figure 3 below illustrates how moderation typically occurs in practice. Moderators first familiarise themselves with the assessment and marking criteria. Next, moderators are standardised to bring them into line with the standards required of the subject specification, in terms of marking and awarding body procedures. Using predetermined sampling criteria, the moderators then select and re-mark a representative sample of work from each of the schools in their allocation. If the school's and moderator's marks are within a specified tolerance, the school's marks for students' work are accepted unchanged. On the other hand, if the school's marks fall outside of that tolerance, the marks are analysed further using statistical procedures such as regression analysis or by judgement. The school's marks may then be adjusted to bring them within tolerance and in line with the required standard, although the rank ordering of students within the school will remain unchanged. The use of statistics (i.e., regression) in this moderation process is solely to automate what would otherwise be a manual procedure and is not a form of statistical moderation.
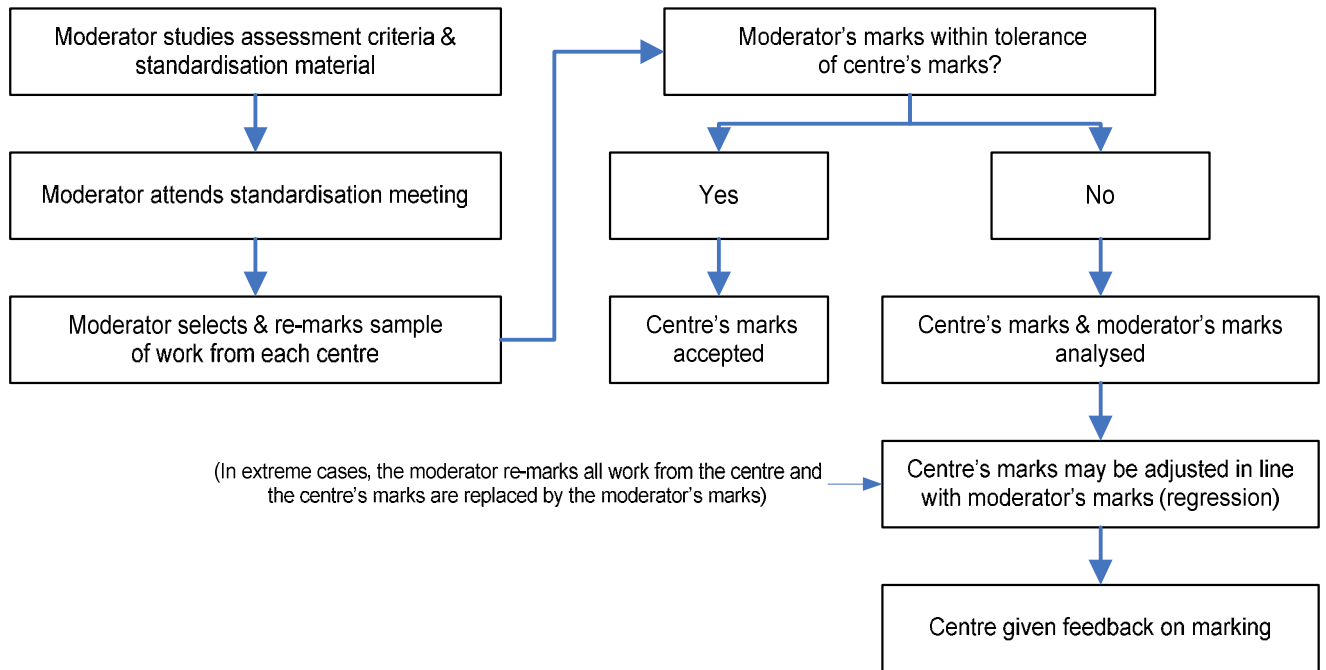
**Figure 3: The AQA moderation process.**

## SUMMARY

This chapter outlined a variety of types of moderation and verification of internally-assessed work. In providing some comparisons between the different forms of moderation and verification, it showed that the moderation of internal assessment is a quality review and assurance process, which aims to ensure that assessment tasks, procedures and practices are valid, reliable and aligned with required standards. Moderation procedures generally seek to foster and encourage good practices in assessment, teaching and learning. It is also reasonable to expect that moderation should help teachers to achieve assessment decisions which are right the first time. These principles provide a basis for pre- or post-assessment moderation to enhance the comparability and quality of assessment and the adherence to standards, thereby leading to improved outcomes that meet expectations.

Anthony Daly
13 October 2009

## AQA RESOURCES

- *Diploma Principal Learning units: Internal assessment and moderation and guidance for centres. September 2008.*
  http://store.aqa.org.uk/admin/crf_pdf/AQA-DIPLOMA-PLU-IA-MOD.PDF

- *New moderator training.*
  http://intranetm/dept/wg-emtsg/2.%20generic%20resources/2.%20face%20to%20face%20training/4.%20additional%20training%20for%20new%20examiners%20&%20new%20moderators/new%20moderators/1.%20training%20for%20new%20moderators.pdf

- *Procedural guidance. Coursework and portfolio units/components in ELC, FCSE, FSMQ, GCE, GCSE, Key Skills, Level 1 & 2 Certificates and Project Qualifications 2009. Moderation procedures.*
  http://intranet/pdf/moderation_pgf.pdf

- *The moderation of centre-assessed units/components. An explanation for centres.*
  http://store.aqa.org.uk/over/stat_pdf/MODERATION.PDF

## ADDITIONAL READING

Australian Council for Educational Research. (2006). *Australian Certificate of Education: Exploring a way forward.* Canberra, ACT: Department of Education, Science and Training, Australian Government. http://www.dest.gov.au/NR/rdonlyres/DEAB7529-F499-448E-B353-E7835DB68180/10319/AustCertificateofEdu2006_WEBPublication.pdf

City & Guilds. (2008). Ensuring quality: Policy and practice for externally verified and moderated assessment. http://www.cityandguilds.com/documents/ind_generic_docs_ensuring-quality/EQ_Mar-08_FINAL1.pdf

Harlen, W. (2007). Criteria for evaluating systems for student assessment. *Studies in Educational Evaluation, 33*, 15-28.

Maxwell, G. (2006, May). *Quality management of school-based assessments: Moderation of teacher judgments.* Paper presented at the 32nd Annual Meeting of the International Association of Educational Assessors, Singapore. http://www.iaea2006.seab.gov.sg/conference/download/papers/Quality%20management%20of%20school-based%20assessments%20-%20Moderation%20of%20teacher%20judgments.pdf

Taylor, M. (2005). Teacher moderation systems. Retrieved October 3, 2008, from Qualifications and Curriculum Authority/Assessment and Qualifications Alliance: http://www.naa.org.uk/libraryAssets/media/Teacher_moderation_systems.pdf

Wilmut, J., & Tuson, J. (2005). *Statistical moderation of teacher assessments: A report to the Qualifications and Curriculum Authority.* www.qca.org.uk/libraryAssets/media/statistical_moderation_of_teacher_assessments_pdf_05_1606.pdf.