

National Foundation for Educational Research

Chartered Institute of Educational Assessors

Policy and research seminar on

Methods for Ensuring Reliability of Teacher Assessments

Tuesday 2 June 2009, at The Royal Institute of British Architects

Proceedings

Edited by Catharine Parkes and Sarah Maughan

Contents

Introduction	3
Presentation 1 Moderation in the International Context.....	4
Summary of presentation	4
Presentation 2 CIEA approach to moderation.....	5
Summary of presentation	5
Presentation 3 Designing systems of teacher based summative assessments	6
Summary of presentation	6
Discussion	7
Purpose of assessment	7
Progression and standards.....	7
Teacher judgements	8
Accountability and Moderation	9
Independence of Teacher Assessment and Testing	9
Changing the system	10
Concluding Remarks.....	10
Appendix 1: List of participants	11
Appendix 2: Presentation by Sarah Maughan.....	12
Appendix 3: Presentation by Graham Herbert.....	17
Appendix 4: Presentation by Richard Daugherty	23

Introduction

The assessment system in England is going through a period of unprecedented change, with changes currently underway or being discussed at every stage of education. There does not appear to be a consistency across these changes, so coursework (one form of teacher assessment) is being replaced by controlled assessment at GCSE, whereas a key component of the new Diplomas will be assessed by teachers. At the lower levels, external tests were removed at key stage 1 in 2005, and key stage 3 tests were abolished from 2009. More recently there has been an announcement that key stage tests in science will be removed from key stage 2 from 2010.

Testing at key stage 1 has been replaced by moderated teacher assessment (although the tests continue to be used to inform the teacher judgements) and there has been some suggestion that teacher assessment will replace testing at key stage 3 and in science at key stage 2. Therefore, although not consistent across the whole system, there does appear to be a general move in the direction of more teacher assessment in schools in England.

This seminar was set up by the National Foundation for Educational Research (NFER) and the Chartered Institute for Educational Assessors (CIEA) to stimulate debate about the best means of building confidence in teacher assessment. In particular the seminar sought to question:

- the most appropriate models of quality assurance of teacher judgements for use in schools in England;
- whether one model could work for all stages of education and for all on-going developments, or whether a range of methods will be required;
- what structures, systems and support need to be put in place before teacher assessment is a trusted part of the education context in England.

Contributions were made by Sarah Maughan (NFER), Graham Herbert (CIEA) and Richard Daugherty (Cardiff University). The seminar was steered by David Wright (CIEA) and Chris Whetton (NFER).

This report summarises the presentations and discussions. The seminar operated under the Chatham House rule, but the three presenters have given permission for their views to be attributed.

Presentation 1

Moderation in the International Context

Sarah Maughan, NFER

This presentation appears in full as Appendix 2.

Summary of presentation

Sarah's presentation aimed to provide a brief overview of the methods that are currently in use around the world to support teacher assessment. She first started by suggesting that the purpose of these methods is, on the one hand, to ensure that the judgements are as reliable as possible, but on the other hand, and perhaps more importantly in some cases, to ensure that stakeholders have trust in the teacher judgements.

Sarah described how quality assurance can be used during the process, in the form of teacher professional development and provision of exemplar assessment materials. However, she suggested that this is not sufficient if the final judgement is not moderated in some way.

Statistical moderation involves adjusting the level and spread of teacher assessments to match that from an external test, or using test results to trigger further checks. In most cases, the rank order of pupils is kept the same.

Social moderation methods include:

- postal – a sample of materials from each centre or teacher is sent to a moderator, who will make recommendations on any adjustments required;
- visiting – a moderator will visit different centres;
- moderation panels – groups of teachers looking at materials from centres (can become people intensive);
- (in school moderation);
- agreement trials – teachers come together to discuss the standard required;
- consensus moderation – similar to agreement trials but take place after the assessments and use live work.

Most moderation systems use a combination of these methods.

There is the question of what is being measured by the teacher assessments, is it the same as tests or is it broader? What can be used as a yardstick to measure reliability and how reliable is good enough? Is teacher assessment less expensive than testing, once all the costs of moderation are included?

Presentation 2

CIEA approach to moderation

Graham Herbert, CIEA

This presentation appears in full as Appendix 3.

Summary of presentation

Quality assurance (QA) is important in teacher assessment to ensure confidence in the system and to ensure the system is fit for purpose. The translation of the principles of QA into systems of teacher assessment is at the very heart of what CIEA are attempting to do.

Applying a QA approach to assessment first brings in the idea of a 'customer' and this customer may be different for assessments with different purposes. Agreeing the purpose(s) of any assessment is key to the system, which can then make it clear who the customers are and what they need assessment outcomes for - there can be many customers who want either individual or collective results for their own purposes. Graham stated that the learner has to be at the centre, with teaching, learning and assessment key to their progression. At the CIEA the QA process is defined by their professional framework.

The procedures include:

- preparing for the assessment;
- conducting the assessment;
- feedback to the pupils;
- reflecting either on learning or teaching.

In teacher assessment, the purpose of the assessment is primarily for the learner. In external summative forms of assessment, one of the main purposes is to meet external needs and so the customer changes. The results are important to schools and governments – they need different information from the tests than the learner.

In forging a QA system for teacher assessment, the process of planning, conducting, moderating, reporting and reflecting must be transparent and able to be scrutinised. Each of these aspects is crucial and no one feature, eg moderation, can be viewed in isolation.

Presentation 3

Designing systems of teacher based summative assessments

Richard Daugherty, Cardiff University

This presentation appears in full as Appendix 4.

Summary of presentation

The purpose (s) of the assessment must be defined in the planning stages and these purposes will be key to determining the nature of the assessment. Individual assessment can be achieved by teacher assessment if results are for those in the school system or even contributing to high stakes assessments such as a record of attainment. Richard suggested that it is not appropriate to use teacher assessment for school based accountability.

The processes involved in teacher assessment should be considered as a series of links in a chain. If any link is weak, the validity of the assessment is compromised. The links or stages that need to be considered are:

- task type;
- task conditions;
- which teachers make the judgements and how;
- criteria and performance standards - need to be clear and to be operationalised;
- how judgements will be used, and for what purposes;
- explicit procedures for each stage;
- teacher expertise - need time and training to develop expertise;
- training and support - needs to be ongoing, not just at the start;
- quality assurance – a moderation process should be chosen that is fit for purpose.

In Wales, summative teacher assessment is being strengthened at all ages. They are introducing a school accreditation process for assessment at the end of key stage 3 and using a school cluster approach for assessment at 11, the age at which pupils transfer to secondary school. At key stage 3, secondary schools become accredited by submitting sample portfolios of pupils' work in all national curriculum subjects for scrutiny and feedback. At key stage 2/3, teachers from secondary and feeder schools come together to arrive at an agreed understanding of the standards achieved in the four core subjects. The support given to schools for this comes from INSET, LAs and Welsh Assembly Government guidance. The process is aimed at building trust; the primary school results become more believable and are used by the secondary schools.

Discussion

The discussion was lively and wide-ranging, covering questions of broad principle as well as practical issues for the short and medium term. The main points raised are brought together here by theme, rather than in the order in which they were raised; where there was a general consensus, this is indicated.

Purpose of assessment

There was a recurring discussion about when and how teacher assessment can be used. There was a general consensus that the appropriateness of teacher assessment comes back to the agreed purpose. If the purpose is closely related to teaching and learning, teacher assessment is appropriate. However, if the purpose is about making judgements on teachers or institutions, there needs to be such stringent moderation practices that the assessment is no longer teacher assessment or it becomes unaffordable.

Establishing the purpose of an assessment is fundamental prior to any other stage of the development as it determines who the results will have value for.

At the moment the purpose of assessment at KS3 is not clear: some feel it marks an end to a broad curriculum entitlement, while it is also used as an accountability measure for secondary schools. While this is open to discussion, it will be difficult to determine the best way of assessing it. Is it a time when pupils are given experience across all the subjects to allow them to make choices and possibly to give some feedback to parents? If it is to track and record national standards, then it needs to be credible and supported. It should be remembered that the majority of subjects at key stage 3 have traditionally been assessed by the teachers, not by external tests, and there has been a general acceptance of this.

There may also be some need to discuss the purpose of GCSE exams. Currently they provide a measure of attainment at the end of statutory education. When the age of statutory education increases, what will be their purpose?

One purpose of end of key stage 2 testing is currently to provide accountability for primary schools. Whilst this is still a primary purpose there may be some difficulty in establishing a system of teacher assessment for this stage.

Progression and standards

A second major theme of the discussion centred around the idea of progression in learning as compared to standards. Does a standards framework provide enough information to allow teachers to chart progression and feed back on next steps in terms of pupil development? Are standards derived from teachers' understanding of progression? This has implications for teacher training.

Setting standards implies a linear progression. However, there is little evidence to imply that progression is linear. One suggestion was to set key milestones at less regular intervals. If the key milestones are set more widely, then the order in which pupils pass the smaller milestones is less important.

Teacher judgements

In discussions surrounding teacher judgements there were two main ideas; how to define levelness and how to establish credibility in a system of teacher assessment.

In testing, levels are defined by cut scores. How do teachers do this on an assessment they have implemented? How do they make decisions about the measurement scale? Some teachers have a holistic view of levelness. With this holistic view, what process do teachers go through when making decisions and what are the issues of transparency concerning this? When the purpose of an assessment is to feed back into teaching then understanding about how pupils have performed on a given task, or set of tasks, is more important than the level that the pupil is working at.

In the early days of the National Curriculum, criterion referenced measures were totted up to give an overview or holistic viewpoint. This approach failed because of its complexity and the variable reliability of the outcomes. In such a system, how does this actually relate to levelness? In a top down system such as has existed in England, not much attention has been paid to feedback to teachers on their assessment skills. Assessing Pupil Progress (APP) aims to re-conceptualise levelness, in terms of progression.

Central to any system of teacher assessment is the recognition that teachers have the expertise to make judgements on pupils' progress. Making any judgement is an internal process and it is intrinsically difficult to justify an internal process as a non-subjective and transparent process. As a result, much of the discussion involved how to establish the credibility of the system of teacher assessment.

It was accepted that, by definition, an expert teacher is also an expert assessor. However, teachers themselves may not believe this, nor may other stakeholders. A key step in the wider acceptance of teacher assessments is to develop confidence in the system. There is a general concern about reliability and bias when using teacher assessment with research evidence about the bias in individual cases. As one delegate pointed out, it only needs one well placed article questioning reliability to undermine confidence. Currently, if a test score and a teacher assessed score are provided, many people would give more credence to the test score. How do you start to change this perception? The test and the teacher assessment may be measuring different things and these need to be recognised.

Ofqual is currently looking at the issue of reliability in assessment, both testing and teacher assessment. This will provide useful data which can be used to assess the relative merits of the two forms of assessment; data which does not currently exist.

When setting up a system of teacher assessment, one way of establishing some credibility is to create clear criteria for assessment. There is a danger of over specifying the criteria, so that it becomes unusable for teachers and markers. There should be some mechanism of having a feedback loop to revisit the criteria over time.

Assessment should become an automatic part of what teachers do; it should be viewed as one of the techniques of teaching, rather than a form filling activity to do at periodic intervals.

Accountability and Moderation

The three main elements of the discussion on accountability and moderation were the need for external moderation for high stakes assessments, the value of moderation and the effects on teachers of increased accountability.

To enable confidence in a system of teacher assessment, there must be transparent checks and balances to show that teachers' judgements are moderated. For high stakes assessments, moderation of teacher assessment is likely to have more credibility if it is external.

There is more pressure on judgements as a result of high stakes assessments. One possible consequence of this is illustrated by some schools in Australia which would prefer not to have teacher assessment, as there is more pressure from parents and pupils to justify the assessment result.

One of the advantages of establishing a system of moderation across schools is the teacher-teacher interaction – this can be very valuable in terms of developing an understanding of the standards, and of the best means of collecting assessment evidence. However, moderation processes that involve teachers extensively can be very expensive. An on-going question is whether this expense is viable in the system or good value for money.

There was some consensus that external assessments are needed where accountability is the main purpose (eg at KS2). This is because of the tension in the teacher's role when they are required to be both the means by which teaching and learning takes place, and the means by which teaching and learning is held to account. In Queensland, one of the most commonly cited contexts in which teacher assessment is used successfully, there is in fact an external test used for creating rank orders for university entrance purposes. Other than this, devolution of decisions regarding assessment has led to a concern that teachers have little experience in national assessment. This has led to some auditing of the school assessment regime across Queensland.

Independence of Teacher Assessment and Testing

There was some discussion about what we aim to do with teacher assessment: are we trying to replicate test results or are we trying to get a broader, or at least different, measure of what pupils can do? Teacher assessment is often judged as reliable if it replicates test results, however, it is also felt that teacher assessment can be more valid because it can assess a wider range of activities than paper based testing can. There was some consensus that they ought to be different things and that different results from the two approaches should be accepted and understood. It was felt that we need to make both assessment approaches independent and valued. One suggestion was that annotated portfolios of work could demonstrate the range of activities being considered when making a teacher assessment.

Where both methods are in use in a system there tends to be a convergence in the results over time. There can be some attempt to get round this by requiring teacher assessment judgements to be made prior to a test being administered. However, over time the

difference between the two still becomes less. There will be a need for some kind of auditing of the two to ensure convergence does not occur, if we adopt a system that aims to actively use both methods.

Changing the system

We need to move to a position in which assessment is an accepted part of a school's 'business'. It should be seen as an accepted part of the professional practice of teachers. This will take a long time from the current position.

To implement a new system, it was suggested that there could be a three year programme of change - e.g. start in year 4 if the final change is to occur in year 6. This would mean that the processes could be developed and embedded in a low stakes environment, so that when they become higher stakes they are trusted by stakeholders and more widely accepted.

Pupils, parents and teachers need to be convinced of the reasons for the change. Some teachers believe that it is a simpler job to use tests and not all currently want to move to teacher assessment. These teachers need to be convinced of the value of the activity.

Concluding Remarks

Concluding remarks were offered by Chris Whetton at the end of the seminar.

He concluded that the key theme to have emerged was that of purpose, and how this needs to be clear before decisions about whether testing or teacher assessment would be most appropriate at different stages of education. How the results will be used and the stakes placed on them need to be agreed. It may be most appropriate for teacher assessment to be used where the primary purpose is to provide information for pupils and parents. Where the purpose is to provide information for use by schools then teacher assessment with some form of moderation may also be appropriate. However, for accountability purposes teacher assessment is likely to need expensive and burdensome moderation procedures, if the results are to be accepted both by the public and the teaching profession. This may make it unviable.


A key issue related to the method to be used for teacher assessment would appear to be the notion of levelness and how teachers are to develop this. There are different requirements of assessment that is used to feed back into learning, and assessment that is used to decide on an overall level. It would be helpful to provide models for teachers of how these requirements can be integrated.

This seminar was set up in an attempt to move forward the debate about possible methods of assessment to be adopted in England. It is hoped that the outcomes from the discussion can be used to inform future decisions, particularly about new assessment approaches at key stage 2 and key stage 3 and the moderation systems needed to support them.

Appendix 1: List of participants

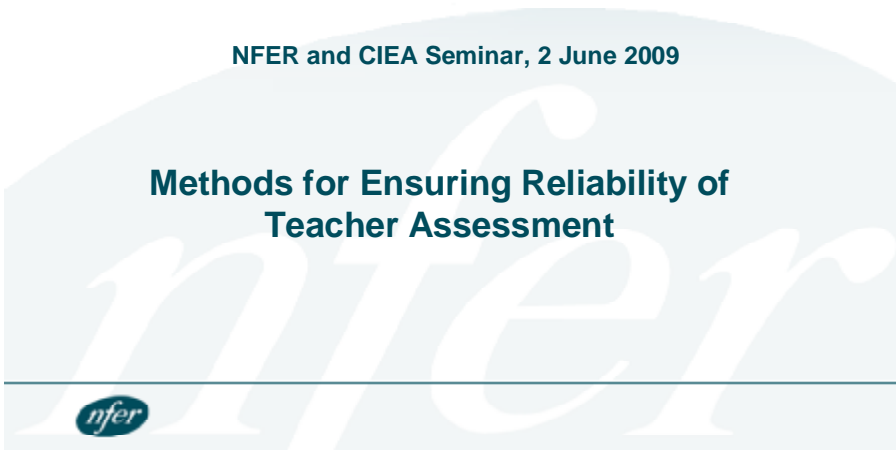
Delegates	Organisation
Chris Whetton	NFER
Sarah Maughan	NFER
David Wright	CIEA
Graham Herbert	CIEA
Richard Daugherty	Cardiff University
Catharine Parkes	NFER
Rob Van Krieken	SQA
Carolyn Hutchinson	Scottish Government
Gordon Stobart	Institute of Education
Peter Tymms	University of Durham
Gordon Stanley	Oxford University
Sylvia Green	Cambridge Assessment
Anthony Daly	AQA
Jeff Goodwin	Edexcel
Sue Horner	QCA
Ali Wood	QCA
Christine Blower	NUT
Kathy Baker	GTCE
John Clay	GTCE
Warwick Mansell	Press

Appendix 2: Presentation by Sarah Maughan



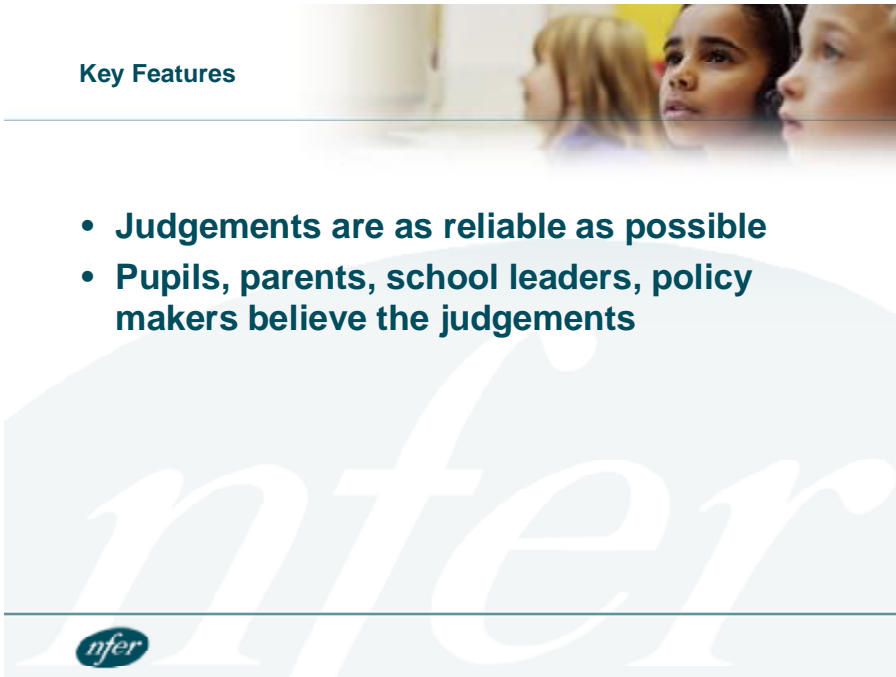
NFER and CIEA Seminar, 2 June 2009

**Methods for Ensuring Reliability of
Teacher Assessment**



Key Features

- Judgements are as reliable as possible
- Pupils, parents, school leaders, policy makers believe the judgements





Quality Assurance

- **At any point during the process:**
 - Design of assessment criteria
 - Design of assessment tasks
 - Delivery of assessment opportunities
 - Judgement of assessment outcomes



During the Process

- **Materials provided to exemplify the kinds of assessments that are to be used**
- **Training provided in the operation of a particular assessment**
- **Steps taken to ensure an understanding of the standards to be applied**





Reliability of the Judgements

- **Statistical moderation**

- Level and spread of teacher marks are adjusted to match the level and spread of marks on an externally set test or examination component
- Or used to trigger further checks



Social moderation: Single Expert

- **Postal moderation**

- A sample of teacher assessed work is sent to a trained moderator for the assessment judgements to be reviewed

- **Visiting moderation**

- A trained moderator visits a centre to review the work and the assessment judgements being made
- May focus on performance or bulky work





Social Moderation: Expert Groups

- **Moderation Panels**

- A group of experienced teachers/ assessors meets to review samples of teacher work and judgements



Social Moderation: Teacher Assessors

- **(In-school moderation)**

- **Agreement Trials**

- Large groups of teachers come together to discuss judgements about samples of work to develop an understanding of the standard

- **Consensus Moderation**

- Similar to agreement trials but after the event discussing 'real' work





Combinations

- In-school moderation
- 'Expert moderation'
- Statistical moderation

- Agreement trials



Summative teacher assessment

- What are we trying to measure?
- How reliable is reliable enough?
- How do we know when we have it right?
- How can we ensure it stays right?
- Can we afford it?



Appendix 3: Presentation by Graham Herbert



The CIEA Approach to Moderation A Quality Assurance Approach

02/06/09



The origins of quality assurance



Quality Assurance

- The CIEA has developed a QA process to help those who assess to implement one in any arena or institution where assessment takes place. It lies at the heart of everything that the CIEA does.
- One popular view of QA sees quality as simply “fitness for purpose.”
- Gold taps are not a sign of quality when all the customer wants are reliable brass ones.
- This popular view also introduces the idea of the **customer**.

Customers

- For any assessment, there can be single customers or a range of customers
- The learner or candidate is one
- The learners’ parents or carers is another
- There are, however, other stakeholders who can legitimately be called “customers”:
 - Institutions in the next phase of learning
 - Employers
 - Local Government
 - Central government
 - Society at large
 - International users of qualifications

Teaching, Learning & Assessing

CIE/ Chartered Institute of
Educational Assessors

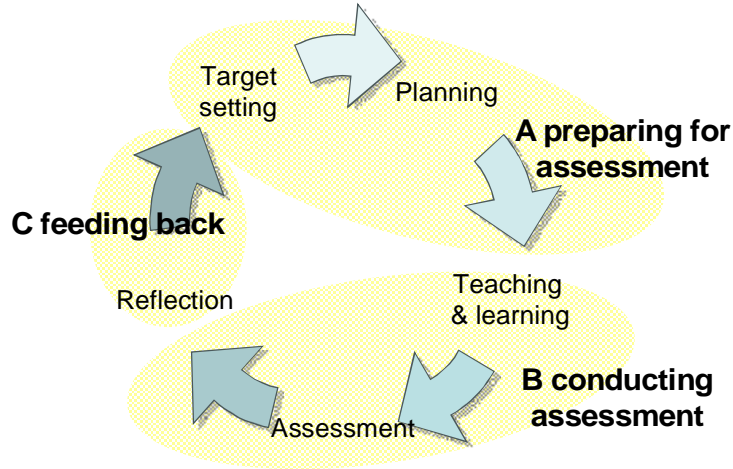


CIE/ Chartered Institute of
Educational Assessors

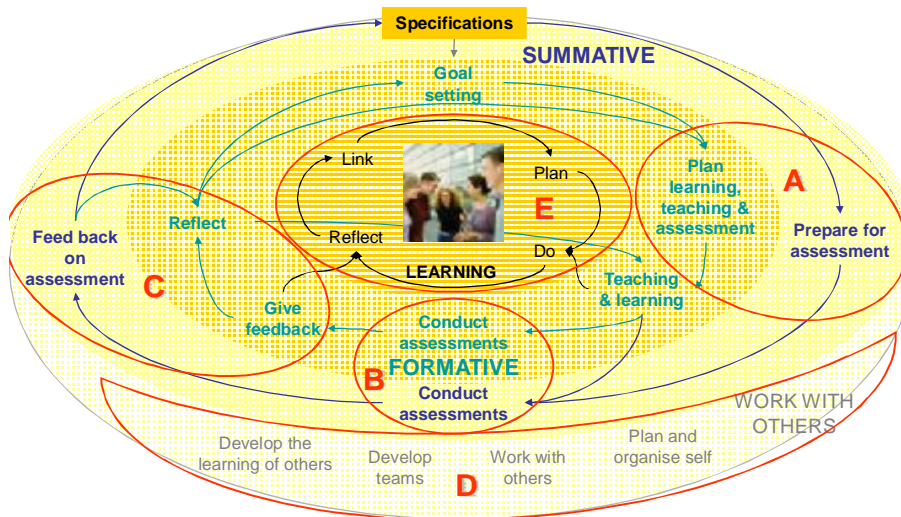
QA Process

- At the CIEA, the QA process is defined by the CIEA Professional Framework
- At its heart is the learner, the primary customer
- The procedures involved in the process are:
 - Preparing for Assessment
 - Conducting Assessment
 - Feeding Back after Assessment
 - The Personal skills needed to function in this environment
 - The Inter-Personal Skills needed to function in this environment

CEI Chartered Institute of Educational Assessors
 Basic Outline of formative framework



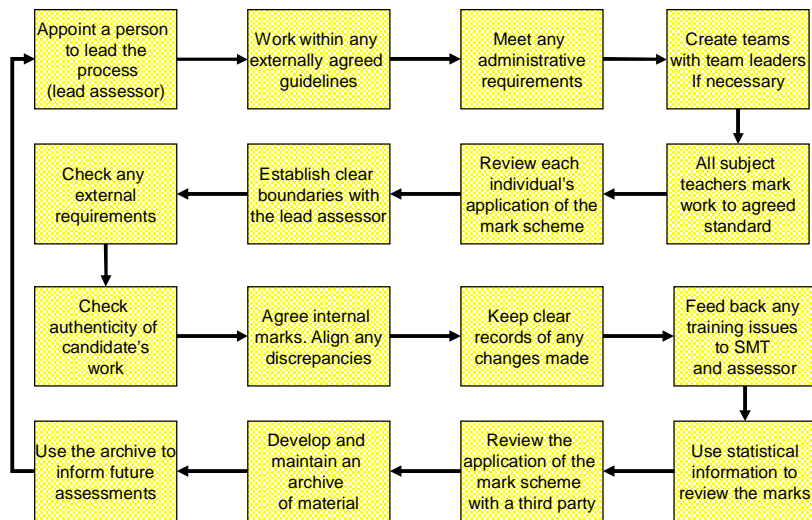
CEI Chartered Institute of Educational Assessors
 Integrated learning and assessment cycle



Moderation

- Moderating assessments would be seen as part of the procedures for conducting assessments with the CIEA QA process
- It is only **one** part of the process and cannot be seen in isolation from the others. It cannot be separated meaningfully from the standardisation procedures.
- However, the moderation procedures can be described in a way that will quality assure the process in whichever context it takes place.

Moderation



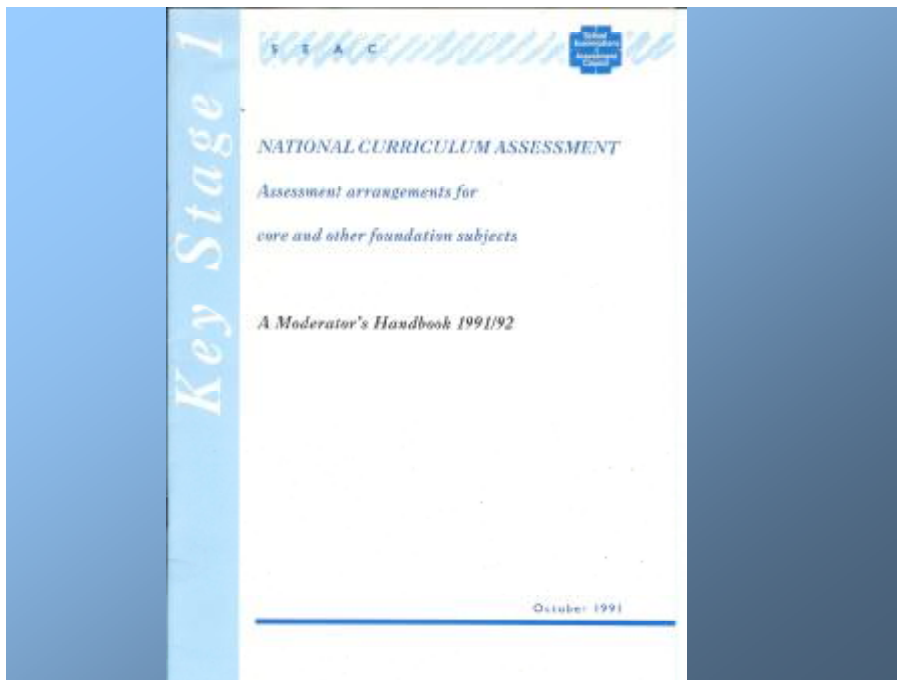
Quality Assurance

- The procedures outlined will provide evidence that each step has been followed to the required standard
- The process will be transparent and open to any external scrutiny including statistical analysis
- That scrutiny could come from a variety of sources: the Local Authority (LA), an awarding body, a Chartered Educational Assessor
- The process will quality assure the assessments, increasing public confidence in the process and satisfying the customer's needs in any circumstance.

Appendix 4: Presentation by Richard Daugherty

Designing systems of teacher-based summative assessment

Richard Daugherty
School of Social Sciences
Cardiff University
daughertyr@cardiff.ac.uk



Systems of teacher-based summative assessment: compatible purposes?

- Performance information about individual pupils for use within or between schools and for parents?
- Performance information on individuals for use as a high stakes record of attainment, including certification?
- Aggregate data as a high stakes indicator for teacher/school/system accountability purposes?

Systems of teacher-based summative assessment: stages in a process

- **Task type** A framework of the type of task that will be assessed.
- **Task conditions** Finding a balance between close specification and unmanageable diversity.
- **Which teachers?** Clarify who will be making judgments and how.
- **Criteria and performance standards** Both need to be clear enough for teachers to operationalise them
- **What inferences will be drawn?** How teacher judgment will contribute to system purpose.

Systems of teacher-based summative assessment: structures

- ***Explicit procedures for each stage*** Pre-judgment and post-judgment procedures spelled out.
- ***Teacher expertise*** Taking account of the expertise teachers bring to the process.
- ***Ongoing training and support*** NOT front-loaded prior to implementation.
- ***Quality assurance and control*** Choose a moderation model that is fit for purpose and ensure that each layer in its operation is effective.

Wales: assessment policies 7 to 14

- ***Developing Thinking and Assessment for Learning*** programme for all age groups.
- ***Cluster group model for moderation***, bringing primary and secondary teachers together to develop consistent standards in reporting of attainment in the four NC core subjects at age 11.
- ***School accreditation model for moderation***, bringing subject teachers in secondary schools together to develop consistent standards in reporting of attainment in all twelve NC subjects at age 14.
- ***Skills profiling***, in terms of number, communication and thinking, to track skills development.

Wales: cluster group moderation model for KS2/3 transition

- ***Teachers involved*** Every Year 6 teacher plus selected secondary subject teachers in the four core subjects.
- ***Main purpose*** To facilitate the transition of each pupil from primary to secondary school.
- ***Moderation model*** Catchment based clusters of a secondary school and feeder primaries.
- ***Support structures*** Designated INSET days, LEA support, WAG guidance materials.
- ***Educational context*** Statutory requirement for every school to have a transition plan in place, including arrangements for end-of-KS2 moderation.