

NCEA Internal Assessment: A harder job than professional marking!

A combined paper from the Manawatu-Whanganui & Auckland Regions

Marking and giving feedback as part of teaching is the professional responsibility of every teacher. No suggestion is being made that this will ever change. However, since the introduction of NCEA in 2002 the administrative and bureaucratic requirements of marking internally assessed standards have dramatically increased teacher workload. The increased burden of internal NCEA marking is due to the NZQA relying on teacher goodwill to develop high quality assessments, mark them, moderate them, and then quality assure them to meet the internal component of the NZQA qualification, all at no cost to the NZQA. There is no incentive for NZQA to reduce this burden, and clear incentives for NZQA to devolve assessment responsibility to classroom teachers. The current proposal for “*common assessment tasks (CAT)*” is an example of this. With a CAT teachers can be sent a formal written exam to supervise and mark on a day selected by NZQA. This is essentially an external paper supervised and marked for free. The timing will then constrain teaching programmes. The teachers who use the CAT will be open to criticism for the quality of their supervision and marking.

In 2009 5,000,000 standards were marked of which just over 3,500,000 were internally marked (NZQA, 2010a). Despite not paying for internal marking and related tasks the NZQA does charge a fee to parents of \$75 per student (NZQA, 2010b). In 2009 the year 11 to 13 roll was 153,844 (NZQA, 2010c). This means revenue from NCEA fees potentially can amount to \$11.5 million. Although NZQA charges a commercial fee for NCEA it contributes almost nothing to the costs of internal marking.

The curriculum realignment is about to place further burdens on teachers with a significant increase to both the amount of internal assessment demanded and the amount of external moderation required for compliance. Although the NZQA assures teachers that everything will be ready and in place in plenty of time, experience has taught teachers that the NZQA generally over promises and under delivers. For some teachers the standards realignment is a much larger task than the original implementation of NCEA in 2002 -2004 with significant increases in the internally assessed component of many subjects.

The NZQA is not the teacher’s employer. However, NZQA acts as if it is an employer of teachers by putting more and more administrative demands onto teachers. Little support has been offered, and often the best support comes through teacher volunteers in subject associations. The quality of assessments provided by NZQA has often been dubious and the range available narrow. The range of exemplars for sufficiency and to delineate grade boundaries has been largely non existent. This has left teachers groping in the dark, learning by trial and error. Schools have been lauded for their performance in a standard one year - then roundly criticised in subsequent moderation.

Assessing for internal and external standards is essentially the same thing, but externals are produced, set, supervised, marked and moderated by paid employees of the NZQA. Assessment of internals is done by teachers who are given little support, increasing compliance activities and are criticised rather than helped or guided if their best efforts are judged insufficient. Marking for internally assessed qualifications to the same standard as external exams, associated administration tasks and the issuing of final grades is different from the normal professional marking duties of a teacher. NZQA must be made to contribute to the costs incurred by schools in these activities and to limit the growing burden they impose.

The NZQA randomised, and increased, the amount of moderation to 10% of internally assessed standards in 2009 to improve the quality of the internal component of its qualification. However, the NZQA has its own quality control difficulties. On August 19 2008 a headline appeared in the NZ Herald "Thousands of NCEA papers wrongly graded" (See Appendix A). Therefore, if NZQA cannot perfect its own marking when it has tight control over the process with several layers of paid experts, how can it expect teachers in 350 high schools to be consistent with marking?

This year NZQA pilot a new system of marking exams in some standards using a score based system and splitting NCEA grades into nine grades (0, N1, N2, A3, A4, M5, M6, E7, E8) instead of the current four (N, A, M, E). Although the grades will not be reported on the student's record of learning they will be recorded on the exam paper. The reasons for the changes are simple - "the NZQA trials showed greater accuracy in assigning grades using score-based marking" In addition grade boundaries will now be determined by cut score. The cuts scores will be decided at the end of the marking process after markers are familiar with student answers. Internal markers do not have this luxury, and must ensure their schedules are perfect to start with (for more information see appendix B).

The message the NZQA seems to be saying after nearly a decade of NCEA is we haven't got it right yet but we expect teachers, many of whom are forced to work in relative isolation, to be perfect. So the NZQA makes compliance in schools more and more difficult and it expects teachers and schools to comply for free to make its qualification look better.

Teachers using internally assessed standards are expected to:

- Set and supervise assessments
- Write or adapt assessments to meet the national standard
- Generate assessment schedules to meet the national standard
- Have assessments pre moderated
- Have marking internally moderated and cross checked
- Allow resubmission and reassessment opportunities for students

- Moderate other teachers work
- Maintain banks of files for external random moderation
- Record grades
- Respond to external moderation reports
- Issue a final grade to students for qualifications
- Undertake professional development to ensure they fully understand the national criteria
- Maintain annotated benchmark samples of student work.

The role of an external marker is similar (see Appendix C). However, there are key differences. The external marker:

- Does not prepare the assessment (this is a separate position)
- Does not adapt assessments to bring them in line with the national standard
- Does not deal with students personally when they fail or appeal a grade.
- Does not have to maintain files of student work for moderation.
- When the marking is done it is over and papers are returned to students
- Gets paid in addition to their normal teaching position
- Does not normally mark more than one standard
- Is not subject to difficulty in their teaching position if the marking is deemed too variable, they simply lose their marking contract.

Therefore, the job of external marker (see appendix C) is many ways easier than that of an internal marker and yet they are paid per marked paper. External markers mark hundreds of papers, mark one standard and get timely advice and guidance as to how to mark and do so only at one time of the year.

Internal assessment now is different to pre NCEA days. Prior to the introduction of NCEA teachers did mark for internal assessment. The essential differences were that:

- 1) The assessments were not subject to an onerous moderation system.
- 2) Teachers did not issue final grades for a qualification.

Teachers now issue passes for a national qualification from their desk whereas, in the past, they did not. In the past the grades were either issued by the NZQA after scaling or by the school in the following ways:

In School Certificate the internal assessment component was typically by completion of assignments. The teacher prepared the assignment and submitted grades to the NZQA at the end of the year. There was little interest by the NZQA in the nature of the assignments as all grades were moderated by the final exam. The NZQA issued the final grade in the School Certificate results notice.

In Sixth Form Certificate all internal marks from a range of school tests, assignments and school exams were combined. A single, final mark was generated and the final grade was issued by the school. Results were moderated by the cohorts' School Certificate results from the previous year. The school issued the final grade and provided these to the NZQA.

Bursary was completed on a similar basis as School Certificate. Internal results were moderated by the external exam results. The final grade was issued by the NZQA.


This paper does not argue that normal marking tasks are not the job of teachers. It argues that designing assessments to a quality standard equal to external exams, marking them and then issuing final grades in unit and achievement standards is not the routine, unrewarded job of a teacher. This is the role of the NZQA. This is a job to be largely carried out by paid contractors or employees of the NZQA.

This paper also asserts that marking for external achievement standards and internal achievement standards are essentially one and the same. However, in one situation the markers are paid and the other they are not. In addition to merely marking, internal assessors have a range of other responsibilities such as assessment writing, moderation and these functions need to be paid for as well.

This is now the time to consider altering the funding system and the support given to internal assessment by NZQA. Schools desperately need:

- 1) A wider range of pre-moderated tasks for each internal assessment replete with clear judgement statements and exemplars to clarify grade boundary decisions;
- 2) More support for moderation, particularly in small and isolated departments;
- 3) A system to help teachers who are to overloaded to cope with the internal assessment workloads created;
- 4) Mechanisms to share best practice and quality assessments between schools at low or no cost.

Recommendations

1. That the report be received.
2. That PPTA immediately start negotiating with the NZQA to create better support and funding of all internal NCEA marking.
3. That PPTA immediately start negotiating with the NZQA to reduce the workload created by internal assessment, or fund schools adequately for it instead of merely passing on the costs. 



References

New Zealand Qualifications Authority (2010a), Annual Report on NCEA and New Zealand Scholarship. Data and Statistics (2009) May 2010

New Zealand Qualifications Authority (2010b), Fees to parents, retrieved from world wide web on 17/6/2010 at <http://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/entry-into-ncea/fees-for-ncea>

New Zealand Qualifications Authority (2010c), (Annual Report on NCEA and New Zealand Scholarship. Data and Statistics (2009) May 2010



Appendix A

Thousands of NCEA papers wrongly graded – report

NZ Herald

8:21 AM Tuesday Aug 19, 2008

Thousands of NCEA papers were wrongly graded by officials in the last examination round, it has been reported.

Most of the mistakes were through marking errors or processing mistakes, *The Dominion Post* said.

The newspaper said thousands of students submitted their answer booklets for rechecking and about a third - 2033 - were "successful".

Affected pupils' grades, whose original marks were either wrong or had been tallied or processed incorrectly, were then marked up.

About 55 students also successfully challenged the top scholarship results.

Qualifications Authority chief executive Karen Poutasi said the number of rechecking requests was a tiny proportion of last year's 1.9 million exam papers.

NCEA was one of the world's few systems in which answer booklets were returned to candidates. The review process aimed to ensure transparency and fairness.

- NZPA

Appendix B (Memo converted to word)

1 of 3 pages

TO: Principal

Principal's Nominee

All Teachers

Marking External Assessment (MEA) Pilot

In the 2010 examination round, a small number of level 1 standards will be marked using a standards-based scoring method for collecting and collating evidence. Candidates will continue to receive an N, A, M, or E grade on their results notices. However, their examination booklet will show the scores used in assigning that grade, with the final score shown on the front cover. The judgement statement will show the cut-score for each grade and enable candidates to check their result.

The methodology used in the pilot has been endorsed by NZQA's external advisory group, the Technical Overview Group Assessment (TOGA). Members of this group include Professors Gary Hawke (Chair), Terry Crooks, Cedric Hall, John Hattie and Geoff Smith, and Associate Professor Alison Gilmore.

Background

Item response theory (IRT) has been used in NZQA since 2006 across all subjects to analyse the effects of question quality on candidate performance in examinations. As a result of this research, improvements have been made to the format of questions, beginning with a move in 2008 to scaffolded items, each having a single Not Achieved, Achieved, Merit or Excellence grade. An item may be a single question or several related questions. All items provide opportunities to reach Achieved, Merit and Excellence. This change was incorporated in the majority of papers in 2009, with the remainder expected to incorporate the change in 2010. Alongside this research, a project was set up to consider using scores to assign grades. Initially, scores were assigned to N, A, M and E to arrive at an appropriate grade. This rough estimate showed that the score-based method may be a more reliable way of assigning grades than the current sufficiency method and full trials were begun in 2007.

The assessment schedules were written with scores assigned to nine levels of performance, based on the criteria in the standard, as follows:

No evidence	Not Achieved		Achieved		Merit		Excellence	
0	N1	N2	A3	A4	M5	M6	E7	E8

The numbers correspond to levels of evidence as shown in the assessment schedule e.g. M6 corresponded to evidence for high Merit performance. Marking panels working in March/April 2008 and 2009 re-marked copied examination papers using the score-based schedules. The original papers were unaffected and returned to candidates as usual.

Benchmarking by the panel leader and senior marker was used to set the appropriate level of candidate response for each score for an item using 'guinea pig' papers. The scores across all items in a paper were then aggregated. Towards the end of marking, the panel leader and senior marker then held a standard-checking meeting, called the cut-score meeting, where samples of candidate work were used to make a final judgement of performance against the standard. The cut-score meeting is an additional stage of benchmarking which is very effective because it takes place late in the marking process when the markers are very familiar with the student answers.

These trials showed greater accuracy in assigning grades using score-based marking. Surveys of the experienced markers involved indicated that once used to the system, markers preferred the score-based system and found it straightforward to use.

Marking External Assessment (MEA) Pilot 2010

The marking pilot this year will be a live pilot. The standards involved will all be at level 1 and have been selected because they cover a range of types of subjects and cohort sizes. The MEA pilot will use information gathered from the previous trials and will use a schedule with the same nine levels of performance.

Subject Standard(s)

Accounting 90026
Biology 90163
90167
90168
Dance 90005
Economics 90197
90198
English 90057
French 90087
History 90214
Japanese 90105
Mathematics 90153

The MEA marking process will have five key differences from the present marking process:

- the collection of evidence is score-based
- scores are aggregated to give a total score across a paper
- markers record the aggregate score for each candidate
- the cut-score phase is the final standard checking phase
- when the cut-scores are entered the grade is generated automatically from the database.



The candidate results notice will not be affected by this marking change.

The marking on the returned papers will, however, appear different:

- each item will have an overall grade and score, e.g. a high M will be an M6, a low A will be an A3
- the aggregate score will appear on the front of the paper but not the grade
- the grade will appear on the candidate's results notice.

The judgement statement, as published on the NZQA website in January, will show the cut score for each grade boundary. This will enable candidates to check their grade has been

generated correctly.

Benefits expected from score-based marking include:

- closer alignment of marking to the standard
- greater accuracy at grade boundaries
- reduction in year-by-year variability
- increased transparency.

Enquiries

Please refer any enquiries relating to this circular to your School Relationship Manager.

School Relationship Manager

Secondary Assessment and Liaison

Telephone: 04 463 3000

Fax: 04 463 3113

Email: firstname.lastname@nzqa.govt.nz

Circulars are available from the New Zealand Qualifications Authority website:

www.nzqa.govt.nz/publications/circulars

Questions and answers about the MEA Pilot

Q1. Why is NZQA carrying out this pilot?

Research over the past four years has been accepted by the external NZQA Technical Advisory

Group¹ as evidence for the value of score-based marking to collect evidence for standards-based assessment.

Q2. Is this a move away from standards-based assessment?

No. This is a change in the way the evidence is collected for standards-based assessment. There will be two benchmarking stages at which the standard will be the basis for decision-making. The first will be when the individual items (or questions) are benchmarked against the requirements of the standard; the second will be when the cut-scores are set for each grade boundary, based on the requirements of the standard. Two benchmarking stages, one late in marking, further improves standards-based assessment and is consistent with other NCEA enhancements.

Q3. Why the change from grades to scores? There is no change from grades to scores – students' final results will continue to be grades.

Research has shown that aggregating item grades by totalling scores is a more reliable way of reaching a grade than the present system. Totalling letter grades is not possible. Totalling numbers is straightforward. The cut-scores then allow the scores to be converted back to grades. There will be no difference in the way final results are presented.

Q4. Why not just report the scores as percentages? The grades are assigned scores which are totalled to give an overall score for the booklet. These scores are not marks. For example, N2 does not indicate 2 marks out of a possible 8 marks. Rather, it indicates that the answer has almost, but not quite, achieved the standard. A3 would indicate that the student had achieved the standard but has done so at the lower end of the achievement band. The score cannot be validly viewed as a mark because it is merely a measuring tool used to collect the evidence for a grade. It allows evidence to be aggregated in a way that using letter grades alone does not.

Q5. Why not use A, A+ rather than scores? See Q3. The scores represent grades but allow aggregation.

Q6. Why have two scores for 'not achieved'? Surely those students have not reached the standard?

There are two scores for Not Achieved, N1 and N2. This recognises that a student who has not reached the standard may have provided some evidence towards achieving the standard. This can be important in the total score. For example, in an examination with four items, a student with four scores, A4, N2, N2, M6, will have a total score of 14. If the cut-score for Achieved is 12 (remember benchmark scripts and the standard will determine this) then the student would receive an A grade. If the evidence in the Not Achieved answers did not contribute to the total score the student would have a total of 10 and would have received N.

Q7. Can we use this system for marking internal standards?

The score-based system involves two stages of benchmarking and may be more complex than current systems used in schools. Nevertheless schools may wish to use it once they become familiar with it.

Q8. How have the standards been chosen? All the standards are level one. They offer a wide range of subject and question type as well as cohort size. They all have a single criterion. Most of them will remain in the revised matrices following the standards review process.

Q9. Could the pilot disadvantage students? Marking and results will be carefully monitored to ensure students are not disadvantaged by being part of the marking pilot.

Q10. Will students have different standards in the same subject marked differently?

The way the evidence is collated will be different. However, the student will still receive a grade for each item alongside the score and the result notice will still show only the overall grade. Students and teachers will not need to make any changes to teaching and learning as a result of this pilot.

Q11. Will students be able to choose to have their examination script marked using the current system?

No. Scripts will be marked on an assessment schedule developed by the examiners specifically for the score-based marking system. There will not be an alternative schedule. The score-based schedule will be the only one used for reconsiderations.

Q12. Why do this now? NZQA has carried out score-based marking tests on anonymous examination scripts for several years as post-marking trials. The pilot is being run at this time to test the logistics of the score based marking process in real time marking.

Q13. If the score based system is better, why not introduce it for all standards immediately? First we need to trial the logistics of using score based marking. Specifically, we need to look at issues such as providing training and assistance for panel leaders to train markers to apply the new approach; determining if the new approach adds to the time required for marking; introducing the cut-score step; and managing any impacts on the marking schedule.

Q14. How will I find out more about the trial?

The examination papers, assessment schedules and annotated exemplars of candidate work from the pilot will be published on the NZQA website in 2011.

¹ Members of this group include Professors Gary Hawke (Chair), Terry Crooks, Cedric Hall, John Hattie and Geoff Smith, and Associate Professor Alison Gilmore.

Appendix C

NZQA

ROLE DEFINITION: MARKER

1 1BACKGROUND

Each year Markers are appointed and trained by NZQA to assist with the operation of the external assessment process. Markers must ensure valid, fair and consistent assessment judgements are made against the national standard. Each marker is allocated a number of answer booklets for which they are responsible. Panel Leaders will be responsible for the training of Markers on their marking panels

For Technology, Graphics, and Educational for Sustainability (Level 3), Markers will assess portfolios across a range of standards within a level. Some Scholarship Graphics Markers will assess both Scholarship and level 3 portfolios.

2 2KEY RESPONSIBILITIES

Markers are required to:

- assess an allocation of answer booklets within the given timeframe
- ensure marking is consistent with the national standard
- ensure administrative deadlines are met
- prepare a report for NZQA
- Reconsider and/or re-mark answer booklets, as appropriate.

3 3PROFESSIONAL AND PERSONAL QUALITIES

Markers will demonstrate:

- the ability to maintain confidentiality in all aspects of the assessment process
- curriculum knowledge and teaching experience for the appropriate level
- understanding and experience of standards based assessment
- the ability to work with others
- the ability to work to stringent deadlines in an organized manner
- The ability to make judgements consistent with the national standard under the direction of the Panel Leader.

4 4 FEES

The Marker's fee will be based on:

- a marking component – determined by a per answer booklet \$ rate
- an administrative component - \$300.

The specific detail of the per answer booklet \$ rate will be included in the *Official Agreement*.

Note:

- (i) The total payment is dependent on the number of candidate entries in each achievement standard, and the number of answer booklets allocated to each marker. Every effort will be made to give all Markers a full allocation of answer booklets. **However, in some situations, this allocation might not be possible.**
- (ii) The total payment will be subject to alteration for:
 - absentee candidates
 - allocation of additional answer booklets
 - any teacher relief day payment made to schools or other institutions to release the Marker to attend meetings.

5 OPERATIONAL EXPENSES

Certain operational expenses may be reimbursed by NZQA. Further details regarding operational expenses and specific fees are outlined in the *Official Agreement*.