

# **TOMLINSON REVISITED**

A supplement to Occasional Paper No. 1  
produced by

**The Exam on Demand Assessment Advisory Group**

sponsored by  
**Exam on Demand Ltd**

**September 2005**

Exam on Demand Ltd  
Broad Farm Business Park, North Street, Hellingly  
East Sussex BN27 4DU

Series Editor: Dr Michael Kingdon  
Chair: ExoD Assessment Advisory Group



**EXAM ON DEMAND**

## Background

---

The first occasional paper (OP1) produced by the Exam on Demand Assessment Advisory Group (AAG) reviewed the recommendations of the Tomlinson Committee, which had reported in October 2004<sup>i</sup>. It raised serious concerns about the reliability of the assessment systems proposed by the Committee, and the institutional burden that they would create. Regret was also expressed that the recommendations of the Tomlinson Committee made only token reference to the current development of e-assessment systems. Attention was also drawn to the ways that the Committee's recommendations conflicted with the blueprint for the development of e-assessment produced by QCA in April 2004.

OP1 was released to the media on 17 January 2005 and a week later the editor of the series left for a visit to New Zealand. His arrival in Auckland coincided with the release of the first Level 3 and scholarship results for the New Zealand Certificate of Educational Attainment (NCEA). Publication of the scholarship results produced an immediate outcry from schools, students and parents.

The NCEA replaced the New Zealand school certificate examination, which had been in use since national school examinations began. The innovative features of the new system include:

- \* Three main levels, the highest of which (Level 3) is required for university entrance;
- \* A scholarship level, referred to by different commentators as Level 4 or Level 3+, which leads to the award of university bursaries. Examinations are based on the Level 3 courses of study but assess higher learning outcomes. The financial rewards associated with gaining individual scholarship passes are modest. Students need to pass in several subjects and achieve passes above the minimum to accumulate worthwhile contributions to their university costs. However, with loans as the alternative, competition for scholarship awards is fierce;
- \* The use of criterion referencing of the standards of performance required for each subject and the 3 main levels. The local description of the system is that it is "standards-based". The move to criterion referencing resulted from ongoing concerns about standards in New Zealand schools and fears that the school certificate examinations, which involved extensive scaling of marks, might have been obfuscating significant differences in the real standards that were achieved in different subjects;
- \* The use of internal assessment and institutional moderation for the award of Levels 1, 2 and some parts of Level 3. Assessment of the scholarship level, is wholly external by written examinations or the submission of portfolios of work;
- \* A new body, the New Zealand Qualifications Authority (NZQA), was created to oversee the implementation of the NCEA, train the teachers for their roles as assessors and moderators, act as the awarding body and be the regulator for other New Zealand qualifications.

The outcomes of the NCEA Level 3 and scholarship awards are of concern to the readers of OP1. First, the NCEA Level 3 and scholarship results were a disaster. Second, the NCEA assessment structure has many similarities with the one recommended for use in England by the Tomlinson Committee. The failures of the former illustrate - and vindicate - the concerns expressed in OP1 about the latter.

## The first NCEA awards

Initial media attention focussed on the huge differences in the percentages of students that had passed the scholarship examinations in different subjects and a 50% reduction in the overall pass rate from previous years.

Table showing the percentage pass rates for the 2004 NCEA Scholarship examinations  
(Source = Report of the Scholarship Reference Group, March 2005)

Rank	Subject	Pass%	Rank	Subject	Pass%
1	Te Reo (Maori)	75.0	14	Spanish	36.3
2	Chinese	74.0	15	Statistics and Modelling	36.3
3	Visual Arts (P)	64.5	16	History	29.2
4	French	54.5	17	Mathematics with Calculus	24.8
5	Accounting	51.2	18	Technology (P)	20.4
6	German	50.7	19	Graphics (P)	13.3
7	Japanese	44.2	20	Science	10.0
8	Latin	42.1	21	Chemistry	9.0
9	English	38.0	22	Economics	7.9
10	Drama (P)	38.4	23	Physics	3.4
11	Music Studies	37.5	24	Media Studies	1.6
12	Geography	37.2	25	Biology	1.4
13	Classical Studies	37.0	26	Physical Education	0.0

P = Assessment by portfolio instead of written examination

The differences in the results for languages and sciences are immediately obvious. Pass rates for the former varied from 75% for Te Reo ((aori) and 74% for Chinese to 0% for physical education. The lowest percentage of passes awarded for any language was 36.3% for Spanish. The highest for a single science was 9% for chemistry and the lowest was biology at 1.4%. Subjects from the same curriculum area frequently had very different pass rates. While 37.2% passed geography, and 29.3% history, only 7.9% passed economics. The New Zealand media declared the differences to be unacceptable.

Further issues about the standards and administration of the NCEA arose as the media frenzy developed. Instances were also reported of students who had been very successful at Levels 1 and 2 but had failed at Level 3, and therefore were denied university entrance in February. Two students interviewed on television suggested that they had been misled about the expectations of the external examinations for Level 3 by over-generous internal marking of the Level 1 and 2 units. There were also reports of students whose entry records or papers had been lost. The media were also quick to comment on the significantly higher results for examination components that had been internally assessed by over those that had been externally assessed.

Some schools were criticised for their management of internal assessment and record keeping. At least one head teacher interviewed in the media admitted that record keeping in his school had been poor and as a consequence, the school had failed to assess all of the criteria required for the award of Level 3. His response was to invite the students concerned to return to school before the end of the summer holiday to undertake a further assessment that would cover the missing criteria. As reports of similar cases followed, overlooked Level 1 numeracy and Level 2 literacy units, both of which are required for the award of Level 3, were quoted as the main omissions. Universities were left questioning whether they were able to accept students upgraded in this way and there were media

reports that at least one had taken legal advice on the matter. Accusations were also made by individual head teachers that other schools in their areas had allowed students to attempt internal assessments as many times as they needed to pass, while their schools had confined students to the single re-sit specified.

Within a matter of days the list of concerns had become a long one. Not surprisingly, head teacher, teacher, student and employer groups blamed the NCEA and called for it to be scrapped.

## **Political fallout**

At first David Benson-Pope the Associate Minister of Education, who had only taken up his portfolio and responsibility for the qualifications system only a few weeks before, did nothing. A government spokesperson even appeared on television and tried to convince his interviewer that the scholarship results showed that science students had simply not achieved the standards of their peers that had sat the language examinations. However, after two weeks Benson-Pope was compelled to respond and announced a new system of "Government Bursaries" to ameliorate with the perceived unfairnesses in the scholarship results. Two Government enquiries were launched; the first into when Ministers became aware of the problems and the second - a scholarship reference group<sup>ii</sup> - which Benson-Pope chaired. Its brief was to identify the issues that required immediate action to ensure the integrity of the coming year's awards. Its membership contained state school principals, university staff including professors of education and a vice-chancellor, a prominent teacher trainer and a representative of the independent sector. No representatives of NZQA were invited to be members but its chief executive attended.

The group reported on the last day of March and made 26 recommendations that included restatements of the aims and principles of the scholarship examination, rejection of any return to scaling, improved feedback to schools and students, later publication of results to provide further time for marking and quality assurance, new principles for the scholarship awards and new ways for NZQA to monitor the examination. Their recommendations were discussed in the New Zealand cabinet the next day and all but one were accepted.

In what Benson-Pope is reported to have described as "an elegant solution, and one that kept the best of the new and old systems", the New Zealand cabinet agreed to retain the standards-based system but rank future scholarship students for each subject and link the percentages of scholarship passes per subject to the top 2% or 3% of the number that sat the subject at Level 3. A "premier award" was also announced for the top 5 to 10 students in each subject. Equally importantly, some variations in the percentages of passes between subjects and over time were considered to be natural and acceptable.

## **Other concerns**

Throughout the years that the NCEA has been in development and the first cohort of students has been studying its courses, teachers and others have voicing ongoing criticisms that reached crescendos as the first results for each level were published. Initially, the focus of these concerns was the reliability of the internal assessments, and their inter-school comparability. While these worries subsided as the first students completed Levels 1 and 2, and more teachers became familiar with their roles as assessors, the differences between the internal and external assessments for Level 3 raise further comparability issues.

Despite the quality and quantity of information that criterion-referenced systems of assessment can provide, the reporting of performance in individual standards is limited to 'achieved' (A), 'achieved with merit' (M) and 'achieved with excellence' (E). Reporting of some units is restricted to 'achieved'

only and standards that are not achieved go unreported. A graphical method of reporting is being used which has further confused many users. The media have also exploited instances when the NZQA results website, developed to enable students and schools to access their results, has become overloaded at key times.

And bad news continues. First, on 5 April a survey of teachers and pupils conducted by NZQA revealed that 21% of pupils dropped out of their courses leading to individual subject standards.

So worried have some of the New Zealand independent schools been about the NCEA that they have been preparing their students for the qualifications offered by Cambridge International Examinations. Indeed, CIE even opened a branch office in Auckland in June 2002 and this year New Zealand students achieved some of the highest marks in the CIE examinations

## **Comment**

The introduction of the NCEA has represented a huge leap in assessment terms. Before the NCEA was introduced New Zealand was still using the school certificate, an examination that England, Wales and Northern Ireland abandoned in favour of GCE O-level in 1950. To have moved from a school certificate to a criterion referenced system in one step is equivalent in UK terms to moving from 1940s school examinations to a Tomlinson style set of graded diplomas without the experience of the GCE, CSE and GCSE, the various incarnations of AS levels and national curriculum assessments. To undertake such a large change in one go can be described, in civil-service parlance, as "brave"; to have done so without initial trials and pilots is doubly so.

Despite the justifications used, the reintroduction of norm referencing for future scholarship examinations constitutes a rejection of the main principle underpinning the NCEA. Some differences in the percentage pass rates for each subject, and for Level 4 as a whole, must have been expected by NZQA. Robert Wood, who had worked in UK, Jamaica, Australia and New Zealand, warned as early as 1991<sup>iii</sup> that the first casualty of increased use of criterion-referenced assessment in public examinations would be inter-subject comparability. Wood reasoned that the writing of subject performance criteria could easily become so inward looking that characteristics of the subject would be emphasised above realistic expectations about student performance. To have engineered the inter-subject differences reported above suggests that the distortion Wood warned about has occurred in the NCEA to such an extent that the many of the Level 1 to 3 criteria may be invalid.

Exploration of the columns of the New Zealand Herald reveals that, since development began there have been many criticisms of the NCEA, its principles and features. A significant proportion of those criticisms have now been illustrated. Of course, all changes to assessment systems attract valid and the invalid criticisms but in this case the opinions of many informed people do seem to have been ignored.

## NCEA and Tomlinson compared

---

The new NCEA and the Tomlinson recommendations share several key features:

1. A four level system of qualification for the 14-19 age range, that incorporates the existing qualifications for the age range;
2. A move from external assessment, with external moderation of any internally assessed units, to a system based predominately on internal assessment and institutional based moderation;
3. A move from norm to criterion referenced assessments. Although Tomlinson did not expressly advocate such a move and left the principles that were to underpin its assessment system open, a significant move from norm to criterion referencing would be necessary to create the a priori criteria on which the internal assessment and institutional moderation would be based.

The NCEA experience illustrates the concerns in OP1, especially those relating to:

- \* The reduced reliability of internally marked and moderated assessments (OP1, page 7) and the consequent reductions in inter-institutional, inter-subject and inter-year comparabilities (OP1, page 6);
- \* The validity of internal assessments (OP1, page 8);
- \* Manageability and the institutional burden of assessments (OP1, page 6);
- \* Record keeping and reporting (OP1, page 14).

Although the UK Government rejected the Tomlinson Committee's recommendations, the ongoing debate is stressing the coherence and importance of the 14-19 stage of education. Further, the distinctions between 16+ and 18+ methods of teaching and assessment on one hand, and between academic and vocational qualifications on the other, are being shown to be increasingly artificial<sup>iv</sup>.

The NCEA experience provides a valuable case study in qualifications change management. The main lesson for countries contemplating changes to assessment systems is that, if surprises and damage to students' life chances are to be avoided, evaluations of the validity, reliability, practicability and acceptability of all untested features are essential.

**Exam on Demand Assessment Advisory Group  
September 2005**

### Endnotes:

---

- i ExoD (2005A), **A response to the Tomlinson Report by The Exam on Demand Assessment Advisory Group** (Occasional Paper No. 1), Exam on Demand Ltd, Hellingly, UK.
- ii Benson-Pope D (Chair) (2005), **Report of the Scholarship Reference Group**, Office of the Associate Minister of Education, Wellington, NZ, March.
- iii Wood R (1991), **Assessment and testing: a survey of research**, CUP, Cambridge, UK.
- iv ExoD (2005B), **The development of e-assessment: 2004 to 2014**, (Occasional Paper No. 2), Exam on Demand Ltd, Hellingly, UK.