

Waikato Journal of Education

Te Hautaka Mātauranga o Waikato



School of Education
Kura Toi Tangata

The
University
of Waikato
*Te Whare Wānanga
o Waikato*

Volume 5 : 1999

Special Section: Teacher Education

WAIKATO JOURNAL OF EDUCATION

TE HAUTAKA MĀTAURANGA O WAIKATO
VOLUME 5, 1999

Vouchers and the Privatisation of New Zealand Education LIZ ELEY AND JOHN CLARK	3
Standards-based Assessment in English: Take 3 TERRY LOCKE	13
Poem: Untitled PIRKKO MARKULA	32
Qualifications, Critiques, and Reforms: The Rhetoric Surrounding the New Zealand Qualifications Authority IVAN HODGETTS AND DARRIN HODGETTS	33
Poem: Boarder's Leave Ending RACHEL WOOD	50
The Maori Language Science Curriculum in Aotearoa/New Zealand: A Contribution to Sustainable Development MILES BARKER	51
<hr/>	
SPECIAL SECTION: TEACHER EDUCATION AND TEACHING: TRENDS AND POSSIBILITIES	61
Introduction BARBARA HAROLD	61
Reviews of Teacher Education in New Zealand 1950-1998: Continuity, Contexts and Change NOELINE ALCORN	63
Research Trends and Possibilities in Teacher Education CLIVE MCGEE	77
Teachers Talk Back: Educational Theory and Teacher Education SUE MIDDLETON AND HELEN MAY	89
Stories to Live by on the Professional Knowledge Landscape D. JEAN CLANDININ	107
Remapping the Practicum in Teacher Education ROSIE LE CORNU (DOBBINS)	121
<hr/>	

The Effectiveness of a Māori-Focussed Teacher Education Programme FRED KANA	133
Poem: Day is Done TERRY LOCKE	138
Coping with the Dual Crises of Legitimation and Representation in Research RICHARD PRINGLE	139
Poem: Tracks JIM DENISON	151
Bright Future, Five Steps Ahead - Making Ideas Work for New Zealand: A Commentary HOWARD LEE AND GREGORY LEE	153
Ka Ora Kāinga Rua: Finding a Home in the Academy. A study into the Experiences of Four Maori Women in the Academy LAURA HAWKSWORTH AND PARE KANA	165
Neo-Liberalism and Constructions of Democracy: The Impact on Teachers' Work NESTA DEVINE	171
Producing an "Iwi-Meaningful" Doctoral Proposal: A Case Study BELLA TE AKU GRAHAM	181
"Coming to Know" in Teaching Physical Education: Moving Across Cultural Boundaries GEORGE SALTER	189
To the Fullest Extent of his Powers: C.E. Beeby's Life in Education NOELINE ALCORN	199
Poem: Junia DEBORAH FRASER	208
Poem: Tania DEBORAH FRASER	209
Poem: Otorohanga, 1966 DEBORAH FRASER	210
Book Reviews	211
Poem: Let's Hear It TERRY LOCKE	217

STANDARDS-BASED ASSESSMENT IN ENGLISH: TAKE 3

TERRY LOCKE

*Department of Arts and Language Education
University of Waikato*

ABSTRACT *Since the 1980s, when standards-based assessment became an officially recommended form of assessment, first achievement-based and then competency-based assessment (in its unit standards form) have been trialled in English and other subjects. While achievement-based assessment (ABA) had considerable support from teachers and has continued to be used in schools, especially those trialling the English Study Design, it was dispensed with as mode of accountability assessment by the New Zealand Qualifications Authority in 1993. Now, with the shortcomings of unit standards assessment exposed, the question arises as to which model of standards-based assessment will win out as Achievement 2001 kicks off. The choice is still between two models.*

INTRODUCTION

A hawk's-eye traverse over the historical landscape of assessment in English reveals a number of prominent features. One is the ineluctable connection between systems of assessment and classroom practices. Another is the conundrum that regularly faces English teachers which might be expressed thus: aspects of our subject are not easily susceptible to examination, yet *not* to examine runs the risk of having the status of these same aspects downgraded in the eyes of students. Yet another is the official ascendancy of standards-based assessment, or assessment against standardised outcomes, that has been arrived at in this decade.

This article constitutes an analysis of, or a "take" of this on-going story in the New Zealand context as 1999 reaches its midway point. It is a story that continues to address three major questions:

1. What aspects of English should be assessed?
2. How should they be assessed?
3. What impact will resulting assessment practices have on classroom practice?

The immediate context for the 1999 story, of course, has been the Government's announcement in November 1998, of the *Achievement 2001* strategy which has established the National Certificate of Educational Achievement (NCEA) as the main school qualification, available at levels 1-4 on the National Qualifications Framework (NQF), and to be introduced in 2001 and phased in over a period of three years. School Certificate and Bursary examinations have been retained, generating credits at levels 1 and 3 respectively. In addition, both Higher School Certificate and Sixth-Form Certificate will be phased out, with the latter to be replaced by NCEA level 2, a

system of internal assessment supported by nationally set common assessment tasks.

Of significance for this article, *Achievement 2001* appears to have staked out an assessment middle ground by suggesting that for "conventional" secondary school subjects, half of the assessment will be by external examination or by centrally generated common assessment tasks and that graded "achievement standards" will replace unit standards in these subjects. Students' annual results will be reported to show credits obtained through both achievement standards and (for "non-conventional" subjects) unit standards, grades from achievement standards and examination marks.

Currently, along with other subjects, English has a Ministry-appointed "experts" panel which has been charged with the following:

1. Developing a subject matrix for levels 1-4 on the NQF. The matrix consists of an identified set of "achievement standards" or aspects of the subject to be tested, each of which has an allocation of credits and an indication as to whether it should be internally or externally assessed.
2. Developing, in conjunction with the work of other expert panels, the form of assessment practice which the achievement standards will embody.

In respect of the three strands mentioned above, then, the panel will be making decisions on what should be taught and how it should be assessed. In the remainder of this article, I will be suggesting that at this juncture (July, 1999) a contest between assessment models is emerging in the work of the panel. It is a contest that has its origins in the late 1980s and its outcome will have, I further suggest, real implications for the third of my strands - the impact of assessment modes on classroom practice.

Take 1 (a): Achievement-based Assessment

As Lennox (1995) has shown, the 1980s were marked by increasing opposition to the continuing hegemony of norm-referenced national examinations as the sole mode of assessment for secondary school subjects. The Department of Education had gazetted an internally assessed School Certificate option in 1982, but students' grades were still moderated by an externally set and norm-referenced test.

In 1986, after widespread consultation, the Department of Education published a report, *Learning and Achieving*, which recommended, among other things:

1. wholly school-based assessment at forms 5 and 6 and a combination of national examinations and school-based assessment for form 7;
2. an achievement-based curriculum;
3. assessment related to standards expressed at three levels (form 5), four levels (form 6) and five levels (form 7);
4. stand-alone moderation at form 6;
5. course design flexibility at forms 5 and 6;
6. a single certificate at form 7 reporting school and examination results separately.

As a consequence of these recommendations, the Department set up trials of achievement-based assessment in English and other subjects between 1987 and 1989. In its 1980s standards-based form, English was broken down into aspects ("Production", "Reception" and so on), and each aspect was evaluated in terms of a set of grade-related criteria (GRCs). Initially, these were at four levels (1-4) but were later changed to five (1-5).

Elsewhere (Locke, 1997), I have commented on the impact of achievement-based assessment on the practices of English teachers. I suggested that students began to have access to sets of clear indicators of what grades actually meant and that teachers began identifying more attentively the components of general tasks such as reading and writing as they began developing rubrics for grading purposes (p. 98).

However, in 1989, achievement-based assessment, whatever its promises and shortcomings, was overtaken by the Hawke Report, the establishment of NZQA and the National Qualifications Framework, and the adoption of a particular form of standards-based assessment (unit standards) as *the* universal mode of assessment for all post-compulsory learning destined to be registered on the Framework.

As Peddie (1995) notes, the demise of five-grade ABA in terms of the Framework and the Record of Learning was signalled in a circular letter to schools sent by the Chief Executive Officer of NZQA on 17 November 1993. Curiously, however, the Authority was content to allow schools to use achievement-based assessment for "providing learning goals, and for formative assessment in the course of that learning" (p. 178).

Take 2: The Move to Unit Standards

One way of capturing the nature of the transformation from an achievement-based mode of standards-based assessment to a competence-based (unit standards) mode is via a transformation exercise. By way of example, I offer an example of achievement-based assessment from a non-New Zealand context, the International Baccalaureate's Theatre Arts course. I show what happens when a set of marking criteria becomes transformed, step by step, into a notional unit standard; what is gained, what is lost. I then want to compare the resultant unit standard with an English unit standard used in the 1990s trials.

Figure 1 is not too dissimilar to the banded marking schedules that have been used for many years in the marking of Bursary English essays, a reminder that criteria-referencing is also an aspect of norm-referenced systems. Even a cursory glance at the descriptor related to mark band 21-25 reminds one of the supposed limitations of this kind of achievement-based assessment. It is full of language, mainly adjectives, which call on assessors to make judgements on the quality of the student's work: "highly imaginative, adventurous, illuminating and original", "effective", "thorough", "practicable", "thoughtful", "sensitive" and so on. Who is to say that the understandings of different assessors are consistent in interpreting these terms?

External Assessment Criteria: Play Analysis and Interpretation	
Mark Band	Descriptor
21-25	The Play Analysis presentation indicates a highly imaginative adventurous, illuminating and original interpretation of the play and an effective coherence of directorial perspective. The candidate demonstrates a thorough grasp of the text's potential as a dramatic experience and is able to articulate a practicable understanding of the processes involved in bringing that text to life. The presentation shows a thorough understanding and a thoughtful, sensitive interpretation of plot and character and is systematic, consistent, poised and mature. There is evidence of a clear and thorough understanding of the nature, function and inter-relation of the elements of production and of how they are combined to produce desired effects.
16-20	The Play Analysis presentation contains examples of an imaginative interpretation of the play and some coherence of directorial perspective. The candidate demonstrates a clear understanding of the text, is able to explore some inventive ideas for staging, and articulates a practicable approach for a realisation of the text. The presentation shows a clear understanding and sensitive interpretation of plot, character and the connection between them, and there is evidence of systematic analysis and of a clear understanding of the elements of production and of the subtleties of their effects. There is a recognition of how the elements constitute an integral whole.
11-15	The Play Analysis presentation gives some indication of an imaginative interpretation, and an occasional willingness to take risks, but there is a lack of coherence of directorial perspective. The candidate has a genuine response to the text and its potential for dramatic interpretation in production, and has some understanding of the possibilities of alternative ideas. The presentation lacks systematic analysis and interpretation, but there are signs of insight in relation to plot and character and the connection between them. there is evidence of an understanding of all the principal elements of production but it is superficial and unbalanced.
6-10	The Play Analysis presentation contains few signs of an imaginative interpretation of the play and little evidence of having adopted a directorial perspective. the candidate shows an understanding of the text and has some imaginative response to some of the more obviously dramatic sequences. The presentation shows some basic understanding of plot and character and there is superficial comprehension of the connection between them. There is evidence of some grasp of the nature and function of the most basic elements of production, but this is partial and inconsistent.
0-5	The Play Analysis presentation indicates very few or no signs of having interpreted the play imaginatively nor of having treated the play as a plan for production. The candidate demonstrates little understanding of genre, style or themes of the text and has little or no idea of how the play could be realised on stage. The presentation shows very little or no understanding of plot and character and is devoid of interpretation. There is very little evidence of a sense of the nature and function of the elements of production and such references as are made to these are confused and inaccurate.

Figure 1: International Baccalaureate: Achievement-based Marking Schedule

A tempting solution to the “subjectivity” pitfall is to switch attention away from intangible qualities such as “imaginativeness”, “adventurousness”, “originality”, “effectiveness”, “thoroughness” and “sensitivity” and towards product-related outcomes of performance that can be unambiguously identified as either present or absent. The 21-25 band descriptor can be transformed into a set of such outcomes by carefully eliminating all of the judgement terms listed in the last paragraph. The result of such a process is Figure 2, certainly a shorter text but still rather dense and a little forbidding. Already it can be seen that a major shift has occurred. While all those awkward judgement words have gone, so also has the qualitative dimension, that is, the invitation to a marker to state how *good* a candidate is at the set task.

The Play Analysis interprets the play and assumes a directorial perspective. The analysis recognises the text’s potential as a dramatic experience and discusses the processes involved in bringing that text to life. It interprets plot and character, and discusses the nature, function and inter-relation of the elements of production and of how they are combined to produce desired effects.

Figure 2: Judgements Into Outcomes

One way of making the set of outcomes in Figure 2 less forbidding is to break them down into discrete elements. Figure 3 does this by breaking an overall task into a set of smaller components. On the one hand, it can be seen that this step clarifies a set of requirements. On the other hand, it can be viewed as potentially fragmenting a task that should best be thought of holistically. At this juncture, we are still one stop away from our unit-standard destination. To make absolutely sure that the standard is the same for everyone, we need to quantify our demands using such markers as “two”, “one” and “five” (Figure 4). In this way, no one student will be doing *too much* to achieve the standard while another is doing *too little*.

- 1.1 The Play Analysis interprets the play and assumes a directorial perspective.
- 1.2 The analysis recognises the text’s potential as a dramatic experience and discusses the processes involved in bringing that text to life.
- 1.3 Interprets plot and character.
- 1.4 Discusses the nature, function and interrelation of the elements of production and of how they are combined to produce desired effects.

Figure 3: Outcomes Into Performance Criteria

One might argue that such a step brings greater precision to the performance criteria. One might also argue that this step effectively minimises the effort a student needs to make in order to be deemed competent (i.e. gain credit). It should also be noted that the addition of quantitative markers still does not introduce a qualitative dimension. Rather it asks for more of the same.

- 1.1 The Play Analysis interprets the play and assumes a directorial perspective.
- 1.2 The analysis discusses two ways in which the text might be brought to life.
- 1.3 It identifies one main plot idea and one idea related to character motivation and relates these to one or more relevant sections of the play.
- 1.4 Five examples of production elements are identified using appropriate terminology and their effect analysed. The effect of combining two of these elements is discussed.

Figure 4: Quantifying the Performance Criteria

This notional unit standard is seen to share similarities with an English unit standard from the 1990s trials, unit standard (US) 12419 (Read poetic written text closely). (See Figure 5). Like the notional unit standard, this one is expressed in outcomes, the qualitative dimension has been eliminated, the task of reading has been broken up into discrete components, and a minimal level of reading performance has been defined using quantitative markers.

Long before the 1997 Green Paper on the Qualifications Framework was released, there was a widespread belief among educationalists that the unit standard model was unsuited to what the *Achievement 2001* project is now calling "school curriculum subjects" (Elley, 1995, 1996b; Roberts, 1997). The pedagogical and educational concerns identified by Hall (1997, p. 34) were:

- the negative impact on course coherence of separating the specification of standards from curriculum development and course design;
- the failure to acknowledge openly the complex nature of most educational and vocational standards and the difficulty in specifying such standards in an easily interpreted form;
- the failure to recognise the impact of process on outcome and the implications of this for interpreting educational standards;
- the "neo-behaviourist" and reductionist nature of the unit standard model and its unsuitability to most general and professional educational contexts;

- the increasing emphasis on assessment rather than teaching and learning;
- the failure to include a focus on excellence;
- the failure to recognise the significance of content and context in assessment of student work and decisions on credit transfer and the recognition of prior learning.

Unit Standard 12419 (Read poetic written text closely)

Performance Criteria

1.1 At least one main idea in the text is explained with reference to at least one relevant section of the text.

1.2 The significance of a main idea in the text is explained with reference to at least one relevant section of the text.

Range: significance relates to social, historical, cultural, physical, political, or personal contexts.

1.3 Five examples of language features are identified using appropriate terminology, and an effect of each example is analysed.

Range: Language features could include figures of speech, sound devices, choice of words, irony, symbolism, grammatical usage, punctuation.

1.4 A technique used to shape the text is identified and analysed with reference to at least one relevant section of the text.

Range: techniques could include structure, narrative technique.

Figure 5: Unit Standard 12419

With particular respect to English, Locke (1997), compared four assessment regimes - Bursary English, Achievement-based Assessment in English, Unit-Standards Assessment and the 1989 Victorian Certificate of Education English Study Design - in terms of scope, validity, the encouragement of best practice, assessment mix, manageability, moderation and reporting. Building on Elley's earlier work (1993, 1996a), I drew attention to the arbitrariness of the eight-level structure and the flawed nature of the notional development in literacy implied in the English curriculum document's achievement objectives and to the inadequacy of some of its key terminology. I further argued that a "flow-on" effect has occurred as flaws in the curriculum document have become embedded in the English unit standard categories, level discriminations and performance descriptors. I also discussed examples of ways in which the outcome-oriented, quantitative articulation of performance criteria and range

statements can lead to the subversion of good classroom practice and a distortion of subject knowledge. After analysing the English "Close Reading" unit standards, I found myself supporting Elley in questioning whether, given the vagueness of the wording of the performance criteria, they in fact assessed what they said they were assessing.

Take 1(b): The English Study Design

The English Study Design (ESD) (Locke & Hall, 1999) was established by a national project team in 1997 and funded by The University of Waikato. Its aim was to contest the official hegemony of unit standards in senior secondary school English, by designing and trialling a programme of study and assessment regime that incorporated an alternative model of standards-based assessment. The ESD was trialled and evaluated at year 12 in 13 secondary schools in 1998 (Hall, 1999; Locke & Hall, 1999).

While acknowledging the misgivings about standards-based assessment of educationalists such as Tuck (1995) and Elley (1995), the project team committed itself to standards-based assessment in the form of grade-related (achievement-based) band descriptors (generic "marking guides"). Such a commitment recognised the value to student learning, especially in the process of formative assessment, of clearly stated achievement objectives and their place in guiding students in setting for themselves language-related goals.

The grade-related band descriptors were seen as a way of articulating the national English curriculum by collating its upper-level achievement objectives, rationalising its language and relating it to a small but comprehensive set of work requirements. The assessment tasks associated with the work requirements constituting the ESD programme at years 12 and 13 were clearly related to these grade-related band descriptors. Figure 6 is an example of a generic marking guide for Writing. It uses a 10-point grading scale to allow fine discriminations in terms of its 5-level grade descriptors. A student's writing would be graded both for two aspects, or learning outcomes:

1. **Content and Context**

their ability to use writing as a vehicle for imaginative response and the development of coherent thought while developing a critical awareness of the immediate and wider writing context, and;

2. **Conventions of Language**

their ability to develop confidence and competence in employing consciously the language features conventionally associated with a particular genre.

	Content and Context	Conventions of Language
E E+ [1-2]	Incorporates straightforward messages in the production of a text. Has a limited sense of the text's intended audience and purpose. Some evidence of a developing point of view.	Can employ some language features (layout, structure, punctuation, diction and syntax) in a straightforward way. Has a limited sense of how these features function in terms of the chosen genre.
D D+ [3-4]	Incorporates a number of linked ideas in the production of a text, especially at the paragraph level. Has a sense of the text's intended audience and purpose or purposes. Indicates a point of view.	Can employ a range of features (layout, structure, punctuation, diction and syntax) in ways that are appropriate to their function in the chosen genre.
C C+ [5-6]	Incorporates and develops with some coherence a number of main and subordinated ideas in the production of a text. Shapes the purpose of the text to a sense of the intended audience. As part of a developed personal viewpoint, has some awareness of cultural factors affecting the impact on readers of the text produced.	Can deliberately and in a controlled way employ a range of features (layout, structure, punctuation, diction and syntax) in ways that are clearly appropriate to their function in the chosen genre.
B B+ [7-8]	Incorporates and develops coherently and in a controlled and deliberate way a number of main and subordinated ideas in the production of a text. Consciously shapes the purpose of the text to a sense of its intended audience. As part of a clear personal viewpoint, has an awareness of cultural factors affecting the impact on readers of the text produced.	Confidently and competently employs a range of features (layout, structure, punctuation, diction and syntax) in ways that are both effective and clearly appropriate to their function in the chosen genre.
A A+ [9-10]	Incorporates and develops with flair, imagination and coherence, a number of main and subordinated ideas in the production of a text. Has a clear set of purposes and a perceived knowledge of the intended audience which is used to position them in a deliberate way. Has a clearly articulated viewpoint and a critical awareness of cultural factors affecting the impact on readers of the text produced.	Shows confidence, competence and flair in employing a range of features (layout, structure, punctuation, diction and syntax) in ways that are striking, innovative and clearly appropriate to their function in the chosen genre.

Figure 6: Marking Guide: Writing

In addition, the ESD utilised the development of exemplary rubrics, that is, interpretations of the generic marking guides which relate particular level indicators to particular language events or specific testing situations, such as responding to a poem or writing an editorial. Figure 7 shows the way a rubric

might be developed for a particular level of the marking guide shown in Figure 6. The rubric provides a marking schedule for a particular written genre, in this case the expository essay. It interprets the general descriptors of the marking guide for Writing by identifying, as level indicators, the sorts of features a student would need to use “deliberately and in a controlled way”, for example, in order to gain a C or C+ in the “Conventions of Language” aspect of Writing. In practice, of course, a single piece of student writing seldom matches neatly a single set of grade-related, level indicators. In practice, as the project team emphasised, the grade for a piece of work tends to be typified by the level at which *most* level indicators are judged to cluster.

	Conventions of Language: Generic Schedule	Level Indicators
C C+ [5-6]	Can deliberately and in a controlled way employ a range of features (layout, structure, punctuation, diction and syntax) in ways that are clearly appropriate to their function in the chosen genre.	<ul style="list-style-type: none"> • Introduction sets scene, gains attention and introduces and defines the topic. Has a clear, relevant conclusion. • Has clearly marked coherent paragraphing. • Draws on a good range of concrete and abstract diction. • Has a range of appropriate sentence structures. • A sound control of the mechanics of writing.

Figure 7: Rubric: The Expository Essay

Teachers in the 1998 trial schools were asked to respond to six questionnaire items related to assessment. Results from four items related to the standards-based model of assessment trialled in the ESD are tabulated in Figure 8.

Hall (1999) notes that:

. . . on balance, most teachers found the standards-based criteria helpful (item 12), [though] a few teachers were less satisfied. . . . With two exceptions, teachers indicated satisfaction with the 10-point scale (items 13-14) although two of the “positive” teachers felt that the scale might be extended, giving responses of 10-20. . . . Given a choice of assessment system for benefiting student learning, most teachers preferred the ESD (item 15). However, one teacher opted for assessment by external assessment and four teachers opted for a combined external examination/ESD system. No teacher opted for unit standards.

In fact, a number of teacher dissatisfactions with the marking guides and rubrics arose out of questions of wording and the limited number of rubrics available in the *Practical Guide for Course Development and Assessment* (1998). These were addressed in the revised version of this publication for use in the 1999 trial continuation.

Item 12: The standards-based criteria used in the ESD are helpful for assessing student work.

	SA	A	N	D	SD
Frequency (%)	5 (29%)	8 (47%)	3 (18%)	1 (6%)	0 (0%)
Index of satisfaction: 76%; Index of dissatisfaction = 6%					

Item 13: The grading scale (A+, A, . . . E+, E) used in the ESD provides enough scope to discriminate between the different levels of performance of your students.

	SA	A	N	D	SD
Frequency (%)	6 (25%)	9 (53%)	1 (6%)	1 (6%)	0 (0%)
Index of satisfaction: 88%; Index of dissatisfaction = 6%					

(SA = strong agreement, A = agreement, N = neutral, D = disagreement and SD = strong disagreement.)

Item 14: Ideally, how many scale points do you think are enough in order to grade student work? Please note that the current grade scale (A+, A, . . . E+, E) comprises 10 scale points.

Responses fell into three categories:

10-20 scale points:	2 (13%)
10 scale points:	12 (75%)
5 scale points	2 (13%)

Item 15: If you had a choice, which method of assessment do you believe would most benefit student learning?

External examination:	1 (6%)
Assessment against unit standards	0 (0%)
The current ESD	11 (69%)
Other: please specify*	4 (25%)
(* All specified external exam + ESD)	

Figure 8: Teachers' Responses in Trial Schools

Take 3: "Achievement" Standards

An English "experts" panel of 12 was appointed in April, 1999. Ten of them were either from schools that had embraced unit standards or were involved in the unit standards trial. No one with direct involvement in the English Study Design trial of its particular version of standards-based assessment was appointed.

Along with other subject panels, the English team was offered a template developed by the Ministry's Qualifications Development Group (QDG), a group of Ministry and NZQA personnel set up to manage *Achievement 2001*, to guide them in the development of achievement standards once the matrix had been decided on (See Figure 9). (In accordance with the constraints I have imposed on the scope of this article, I will not be dealing with general or subject-specific issues raised by the matrix model.)

<i>Title</i>	Summarises outcome from credit
<i>Level</i>	Framework level
<i>Credit Value</i>	Designated number of credits
<i>Assessment</i>	External or internal
<i>Learning Outcomes</i>	Knowledge, skills and processes that contribute to the key outcomes in the standard
<i>Assessment</i>	Against which assessment judgements are to be made.
<i>Criteria</i>	There will be three levels of criteria for each standard: credit, merit and excellence
<i>Explanatory Notes</i>	References to source documents, definitions, expected content coverage and context. Description of quality and quantity of evidence required to reliably infer that the learner has achieved the required outcomes in the new standard. Guidance on nature of appropriate evidence may also be given.
<i>Moderation Requirements</i>	Reference to the external moderation will be given here.

Figure 9: Achievement Standard Template

As it began its work in May, 1999, the English panel had two standards-based traditions available to it: an achievement-based assessment tradition dating back to the 1980s trials, refined in the ESD project and applied in current ESD trial schools; and a competence-based assessment model developed for the unit standards trials and currently in use in those schools which opted for this form of assessment.

Potentially, it had two pathways to follow in developing its achievement standards:

1. An achievement-based assessment model with grade-related criteria at A (excellence), B (merit) and C (credit) but, in accordance with the Ministry's effective retention of a pass/fail model for the Framework, no criteria for D or E.
2. A model attempting to recast competence-based unit standards by developing merit and excellence steps describable in terms of the sorts of quantitative, "objective", product-related outcomes described earlier. (See Peddie 1995 for a discussion of the difficulties involved in tacking merit and excellence steps on to a competence-based model of assessment.)

The QDG offered the panel two development options for achievement standards prior to beginning its work.

AS 1	
LO 1	Each LO has criteria for achievement. Grades are awarded when certain criteria are met in each LO .
LO 2	
LO 3	

Figure 10: Option 1: Sampling Within Learning Outcomes

AS 3	
LO 1	Each LO has criteria for achievement. Grades defined by subsets of LO achieved, e.g. credit may be awarded for achieving LO 1 and LO 2, merit for LO 1-3, (and excellence for LO 1-4)
LO 2	
LO 3	
LO 4	

Figure 11: Option 2: Sampling Across Learning Outcomes

As will be seen, the first option (Figure 10) lends itself to the development of an achievement-based assessment pathway, while the second option (Figure 11) lends itself to the development of a modified unit standard pathway.

In its first meeting, the English panel completed first drafts for six out of the seven achievement standards it had decided on for its draft matrix. Five of these drafts reflected but deviated slightly from the first option (See Figure 12) and one followed the second option.

Without being party to the discussions, one can only speculate on why the panel should have favoured the first option (Figure 10) in its achievement standard development work. The model at least has a familiar ring for teachers who have been using achievement-based assessment since the 1980s. Qualitative judgement diction re-emerges, in words such as "coherent", "striking", "imaginative" and "fluent". Moreover, a comparison with the formulation of work requirements in the ESD programme shows a considerable degree of convergence with this particular achievement standard version.

<i>Title:</i>	Write poetic text
<i>Level:</i>	One
<i>Credit Value:</i>	Four
<i>Assessment:</i>	Internal
<i>Learning outcomes:</i> Students who are awarded this achievement standard will produce a range of poetic texts which:	
<ul style="list-style-type: none"> • Have undergone the process of drafting, shaping, revising and editing • Include a range of appropriate language features • Use the conventions of (poetic) writing accurately 	
<i>Assessment Criteria:</i>	
<i>Credit:</i>	
<ul style="list-style-type: none"> - Vocabulary suits audience and purpose - Produces simple sentences - Accurate use of language features - Routine words are spelled accurately - Uses basic punctuation consistently - Structured in paragraphs - Personal voice begins to emerge 	
<i>Merit:</i>	
<ul style="list-style-type: none"> - Uses wide vocabulary appropriately - Demonstrates some sentence variety - Language features used imaginatively and create an effect - Demonstrates skill in spelling, grammar and punctuation - Structured as coherent whole - Personal voice is evident 	
<i>Excellence:</i>	
<ul style="list-style-type: none"> - Uses extensive and appropriate vocabulary - Sentences are crafted - Striking, imaginative, and effective use of language features - Demonstrates mastery in spelling, grammar and punctuation - Confident, fluent integrated writing - Personal voice is sustained 	

Figure 12: Draft Achievement Standard Using Grades for Each Learning Outcome

As mentioned above, the panel drafted one achievement standard for the second option (Figure 13). In the aftermath of its first meeting, panel member Mike Fowler produced a paper arguing that this option should be adopted by the panel, providing draft achievement standards that he himself had developed and tested. Part of the rationale for his choice of the second option

was manageability, suggesting that the presence of multiple performance criteria in an ABA system would create difficulties for teachers. Fowler cites teacher experiences from the unit standards trials to support his view. However, teachers trialling a genuine ABA system in the ESD trial registered an index of satisfaction of 100% with respect to manageability in terms of teacher workload (Hall, 1999, p. 5).

Title:	Write poetic text
	Students credited with this achievement standard can produce two pieces of poetic writing.
Level:	One
Credit Value:	Four
Assessment:	Internal
Learning outcomes:	(for each piece) <i>Notes to assessors are in italics.</i>
LO 1:	Ideas are developed. <i>(Ideas could include thoughts, feelings, sensory qualities, the writer's voice.)</i>
LO 2:	Conventions of the chosen form (poetry, prose, drama) are observed. <i>(Conventions could include text organisation and expression – i.e.: use of language, literary devices, style.)</i>
LO 3:	Writing is crafted to publication standard. (Publication standard means that technical accuracy in spelling, punctuation and syntax is sufficient so that only minor alterations to non-recurring errors may be required.)
LO 4:	Ideas and conventions are effectively combined throughout the writing. <i>(Ideas are fully developed in conjunction with fluent, sustained expression.)</i>
<u>Criteria for awarding the achievement standard:</u>	
Credit:	Learning outcomes 1 to 3 are achieved for each piece.
Merit:	Learning outcomes 1 to 4 are achieved for one piece, and outcomes 1 to 3 achieved for the second piece.
Excellence:	Learning outcomes 1 to 4 are achieved for both pieces.

Figure 13: Draft Achievement Standard Using Grades Across Learning Outcomes

Figure 13 is an example of Fowler's use of the second option to produce an achievement standard for "Write Poetic Text".

It is apparent that this particular achievement standard version is really just a competence-based (or unit standards) model with additions. Indeed, Peddie discusses "additional learning outcomes" as one possible approach to adding a merit step to a competency-based system and poses the question, "On

what basis would these additional learning outcomes be selected, and how many additional outcomes would be considered sufficient for merit to be awarded?" (1995, p. 191).

In respect of Peddie's first question, clarification is needed as to why a *combination* of ideas and conventions has been chosen as the merit discriminator, and what this actually means. Why not writing at a higher quality (were it possible to reach agreement on what a "publication standard" means)? Why not ideas that are outstanding, lateral and highly imaginative? If the weak word "effectively" is to be countenanced as a tag in the merit discriminator, why not these other terms? Why not a merit discriminator that goes beyond the mere "observance" of conventions, whatever that means, to an application of conventions at a higher level of technical mastery?

In respect of Peddie's second question, it is clear that a huge burden has been put on the selection and wording of Fowler's merit discriminator. It is this learning outcome alone which dictates the award of merit or excellence. And it poses the additional problem of why it should be excellent to be "effective" in two pieces of work, but meritorious to be "effective" in only one. (Two effectives = excellent; one effective = meritorious.) What about the student who writes brilliantly in only one piece and is clearly a "top" student?

I would argue that the difficulties just identified are germane to the QDG's Option 2 achievement standard model and are not specific to Fowler's particular example. What they suggest is that the basis for awarding merit and excellence in terms of this model is too inadequate to be seriously countenanced.

CONCLUSION

One of the overhead transparencies used by the QDG Project Manager to brief the English experts panel at the start of their first meeting, identified a number of "policy drivers" for *Achievement 2001*. Apart from workload and manageability issues for teachers and assessors and public concern for a "perceived lack of credibility", there is mention made of "sector dissatisfaction with aspects of unit standards", including their "applicability to conceptual learning" (Ministry of Education, 1999).

No overt mention is made of dissatisfaction with the unit standards model on the grounds of its impact on classroom practice (Elley, 1996b; Hall, 1997; Locke, 1997). Yet the credibility of an assessment system will hinge, among other things, on its validity, its support of sound classroom practices, and that it assesses what it says it is assessing. Disregarding the persistence of the discredited term "poetic", Fowler's example of an achievement standard in Figure 11 sets down a sensible set of writing-related outcomes. But the way they interconnect in the business of determining credit, merit and excellence end up distorting the business of writing by establishing a flawed hierarchy, with ideas and conventions somehow privileged over craftsmanship. Such flaws inevitably flow over into classroom practice and student (mis)understanding.

It is clear from the Ministry of Education's briefing documents that the Government is under considerable pressure to get its achievement standard model right. If it does not, it risks the same resistance from teachers and other

educationalists that prompted the rethink on unit standards. Yet, at least from the perspective of July, 1999, the potential remains for subject experts panels, especially those stacked with personnel with a unit standards orientation and guided by the QDG's Option 2, to produce achievement standards that are little more than unit standards with clip-ons. As I see it, that would be a recipe for disaster.

Postscript: August, 1999

In July, 1999, the Ministry of Education sent to English departments throughout New Zealand a package of materials which included a revised matrix of achievement standards, a set of individual achievement standards in draft form (including learning outcomes, assessment criteria, and sample assessment activities and schedules), accompanying notes and a questionnaire. Schools were to have their questionnaires returned to the Ministry by August 6.

	Conventions	Ideas	Style
Credit	Conventions are observed (i.e. accuracy in text organisation, syntax, punctuation and spelling)	Ideas are developed	Style of chosen form is observed and appropriate to purpose
Merit	Conventions are observed	Ideas are fully developed	Style of chosen form is observed and appropriate to purpose
Excellence	Conventions are observed	Ideas are fully developed	Style of chosen form is fluent, sustained and appropriate to purpose: (i.e. flair, confidence, displayed)

Fig 14: Write Poetic Text: June Version

Of pertinence to this paper, the draft achievement standards reflected a marked shift in the English experts panel assessment orientation towards Option 2, the unit standards-based assessment model. Whereas the panel in its May meeting used this option for only one achievement standard, the panel in June, for reasons it alone knows, chose the option for *four* out of the seven suggested achievement standards. Both writing achievement standards used Option 2,

whereas in May these were both developed in terms of Option 1. (This meant, for example, that respondents to the questionnaire had no opportunity to view a writing achievement standard developed according to the Option 1 model.)

"Write Poetic Text" (1.1) was by far the most developed of the draft achievement standards and, in fact, was a modification of the achievement standard developed by Mike Fowler in the paper he circularised to panel members before the June meeting and which has been discussed in this article. It does, however, differ somewhat in terms of the learning outcomes it specifies, as shown in the postscript figure.

Although the wording is a little different, this achievement standards suffers from the same difficulties as Fowler's previous Option 2 version. On what basis has "Ideas" been chosen as the criterion which will determine merit (by being "fully developed" rather than simply "developed")? Again, on what basis is "Style" chosen as the determining criterion for excellence? Once again, the discriminations between the levels under each aspect are either non-existent or crude and arbitrary. No recognition is accorded to the fact that conventions, if that is the right word, can be observed in differing degrees. The word "fully" magically appears to signal merit with respect to ideas, as if "development" is the *only* attribute of ideas that warrants consideration. And then, the words "fluent" and "sustained" appear to magically signal that a piece of writing which has erstwhile been merely observant and appropriate has now become excellent.

In August, 1999, we are no closer to knowing what the outcome of this contest of assessment regimes is going to be with respect to English and other subjects. What we do know is that one of the most crucial shortcomings of the Ministry's July questionnaire to schools was its failure to actually spell out the existence and nature of these two assessment options and to give teachers the opportunity to pass at least a preliminary judgement.

REFERENCES

- Department of Education (1986). *Learning and achieving: Second Report of the Committee of Inquiry into Curriculum, Assessment and Qualifications in Forms 5-7*. Wellington: Learning Media.
- Elley, W. (1993). Curriculum reform: forwards or backwards? *New Zealand Annual Review of Education*, 3, 37-49.
- Elley, W. (1995). What is wrong with standards-based Assessment? In R. Peddie & B. Tuck (Eds.) *Setting the standards: The assessment of competence in national qualifications*, (pp. 78-99). Palmerston North: Dunmore Press.
- Elley, W. (1996a). Curriculum reform: Forwards or backwards? *Delta* 48 (1), 11-18.
- Elley, W. (1996b). Whatever happened to quality? *English in Aotearoa*, 30, September, 21-27.
- English Study Design* (1997). Hamilton: University of Waikato.
- English Study Design: A practical guide for course development and assessment* (1998). Hamilton: University of Waikato.

- Fowler, M. (1999). English achievement standards: awarding credit, merit and excellence. Paper submitted to the *Achievement 2001 English experts' panel*.
- Hall, C. (1997). The National Qualifications Framework Green Paper: What future for the framework? *New Zealand Annual Review of Education* 7, 29-58.
- Hall, C. (1999). *Evaluation of the Year 12 English Study Design Trial*. Wellington: Victoria University.
- Lennox, B. (1995). Advocacy, evolution and learning: School assessment for the National Qualifications Framework. Paper presented at NZARE Conference, Palmerston North. 8 December.
- Locke, T. (1997). An essay on assessment: The English casebook. *Waikato Journal of Education* 3, 93-117.
- Locke, T. & Hall, C. (1999). The 1998 Year 12 English Study Design Trial: A standards-based alternative to unit standards. *New Zealand Annual Review of Education* 8, 167-189.
- Ministry of Education (1999). *Qualifications development: NCEA*. Set of overhead transparencies used as briefing papers for subject experts panels.
- Peddie, R. (1995). Competency, standards, merit and excellence. In R. Peddie & B. Tuck (Eds.) *Setting the standards: The assessment of competence in national qualifications* (pp. 177-199). Palmerston North: Dunmore Press.
- Roberts, P. (1997). A critique of the NZQA policy reforms. In Mark Olssen & Kay Morris Matthews (Eds.) *Educational policy in New Zealand: the 1990s and beyond* (pp. 162-189). Palmerston North: Dunmore Press.
- Tuck, B. (1995). Issues of objectivity in assessment: A plea for moderation. In R. Peddie & B. Tuck (Eds.) *Setting the standards: The assessment of competence in national qualifications* (pp. 59-78). Palmerston North: Dunmore Press.