

[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]

The 1998 Year 12 English Study Design Trial: A Standards-Based Alternative to Unit Standards

TERRY LOCKE AND CEDRIC HALL

Abstract:

This paper provides an outline of the development and trialling during 1998 of the English Study Design (ESD), a standards-based programme in Year 12 English adopted by 13 secondary schools in Auckland, Christchurch and Otago-Southland. The programme was developed as an alternative to the unit standard system which the design team considered to be flawed on both pedagogical and administrative grounds. This paper reports on the structure and organisation of the ESD programme, its design philosophy, the assessment procedures employed, the results of the evaluation, and the implications of all of these for the Government's Achievement 2001 initiative. The results of the evaluation show a high level of support for the programme by teachers, but variation between schools in student satisfaction ratings. Of particular significance to the Ministry's Qualifications Development Group are the highly favourable ratings (100% satisfaction) in respect of teacher workloads and administrative manageability. The writers conclude by cautioning against the view that a simple modification of unit standards is all that is needed in order to overcome the shortcomings of the existing system.

The release of the Green Paper on national qualifications in June 1997 can be seen in part as an attempt to breathe new life into the original intention of the Education Amendment Act (1990), that is, to establish a "seamless" system of national qualifications which would ensure that all qualifications "...have a purpose and relationship to each other that students and public can understand" (Para 253 (1)(c)). It was also a response to a problem identified by Hall (1998) as "...NZQA's model of standard-setting and the polarisation of attitudes that has

168 *Terry Locke and Cedric Hall*

developed towards it" (p. 33). Among groups in opposition to the use of unit standards in general and professional education were traditional tertiary providers and many secondary schools.

To some extent then, the release of Government's White Paper *Achievement 2001* in November 1998 was an exercise in appeasement – an attempt to achieve a middle ground in the quest for a widely endorsed shape for future, senior secondary school qualifications. The first aim of this paper is to discuss the rationale, content, structure and 1998 trial of the Year 12 English Study Design (ESD) project which was, in its own way, also an attempt to achieve a middle ground with respect to assessment in that most contentious of subjects, English. Its second aim is to summarise the independent evaluator's report of the Year 12 trial, and to discuss its implications with respect to some of the questions left unanswered in the *Achievement 2001* documentation currently available.

One of the writers of this paper, Terry Locke, is Coordinator of the ESD project. The other, Cedric Hall, is the independent evaluator whose role it was to gauge the responses of English heads of departments (HODs), teachers and students to the trial of the ESD in their secondary schools.

Problems with Unit Standards

By the time the 1997 Green Paper was released, there was already a widespread belief among educationists that the unit standard model was unsuited to what the 1998 White Paper is now calling "school curriculum subjects" (Churchman & Hall, 1997; Elley, 1995, 1996; Hall, 1994; Irwin, 1994, 1997; Roberts, 1997). The pedagogical and educational concerns identified by Hall (1998) 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.

(Hall, 1998, p. 34)

The administrative concerns similarly identified were:

- the array of cumbersome moderation procedures generated by the model;
- the excessive record-keeping generated by the model and the vast bureaucratic structures being created to manage the development and implementation of the model;
- the devolved costs to users of the NQF;
- the need for greater decentralisation of standard setting for rapid handling of changes in knowledge and innovations in a subject or field (this also has a pedagogical basis);
- the excessive workloads for teachers.

(Hall, 1998, p. 35)

This accumulation of resistance has to be viewed somewhat ironically in the context of teacher advocacy for changes in assessment since at least the mid-1980s. Lennox (1995), among others, documents the widespread teacher dissatisfaction with national examinations (including School Certificate and Bursary) as sole points of summative assessment, and with the dependency of a school's Sixth Form Certificate (SFC) grade-pool on the previous year's School Certificate results. After widespread consultation, the former Department of Education published a report, *Learning and Achieving*, which recommended, among other things:

- wholly school-based assessment at forms 5 and 6 and a combination of national examinations and school-based assessment for form 7;
- an achievement-based curriculum;
- assessment related to standards expressed at three levels (form 5), four levels (form 6) and five levels (form 7);
- stand-alone moderation at form 6;
- course design flexibility at forms 5 and 6;

- a single certificate at form 7 reporting school and examination results separately.
(Department of Education, 1986)

Why did widespread support for standards-based assessment in the 1980s transform itself into polarised debate on unit standards-based assessment in the late 1990s? The position the ESD developers would put forward is that NZQA simply got it wrong in the form of standards-based assessment that they attempted to impose, with little consultation and without a proven research backing. In an ironical twist, their commitment to unit standards effectively strengthened the hand of the school lobby which traditionally favoured national examinations. Moreover, they failed to take into account the special nature of English and the problems the subject was already experiencing in attempting to follow a national curriculum document that was flawed (Elley, 1996a; Locke, 1997).

The Case of Subject English

Not surprisingly, the various general concerns about the unit-standards assessment model mentioned previously apply to English as a subject. Elley (1996b) declared himself unimpressed with English unit standards because:

1. They are not standards at all.
2. They are too fuzzy to clarify what students know and can do.
3. They do not differentiate between students of widely differing levels of achievement.
4. They appear to give credit for completion, and ignore the importance of quality.

(Elley, 1996, p. 24)

In addition, he worried about the requirement that all performance criteria be satisfied if credit was to be earned, about the workload implications for teachers and the difficulties of moderation.

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), he draws attention to the arbitrariness of the eight-level structure and the flawed nature of the notional development in literacy implied in the achievement objectives of the English curriculum document, and to the inadequacy of some of its key terminology. He argues that a “flow-on” effect has occurred as defects in the curriculum document have become embedded in the English unit standard categories, level discriminations and performance descriptors. He also discusses examples of ways in which the outcome-oriented, quantitative articulation of performance criteria and range statements can lead to deviations from good classroom practice and a distorted understanding of the very subject teachers are wanting their students to engage with. Again, after analysing the English “Close Reading” unit standards, he supports Elley in questioning whether, given the vagueness of the wording of the performance criteria, they in fact assess what they say they are assessing.

In various ways, the concerns with unit standards discussed above led to a groundswell of opposition among English teachers in the early months of 1997, the year of the first unit standards trial in English. At least 120 secondary school English departments publicly declared their opposition to unit standards. In 1997, the co-writer of this paper set up a widely representative project team to develop a structure for teaching and assessing English in the senior secondary school that would attempt to establish a middle ground between the “rock” of one-off national examinations and the “hard place” of unit standards-based assessment.

The English Study Design

The ESD (1997) was developed over a period of four months in late 1997. During its gestation period, it drew on the experiences of those who had made achievement-based assessment work successfully in their classrooms, the advocacy for change that was active prior to the establishment of NZQA, and the 1989 version of the Victorian Certificate of Education English Study Design. In addition, it drew on and represented a convergence of the competing philosophies and practices of the different members of the project team itself.

What the ESD provided was:

- an articulation of the national English curriculum (which removed some of its deficiencies, as identified by Elley and others);
- pointers on English programme design and classroom practice;
- a revised approach to assessment.

Central to its design structure was:

- a view of years 12 and 13 as a learning continuum;
- a flexible yet comprehensive programme of study described in terms of a set of work requirements;
- a set of common assessment tasks (CATs) for both years 12 and 13;
- a combination of internal and external assessment;
- the use of standards-based assessment using grade-related band descriptors;
- certification in the form of a summary profile at both years 12 and 13.

At year 12, for example, common assessment tasks (weightings in brackets) were a Workfile (10%), a Writing Folio (25%), Response to Text (35%), a Language Investigation Project (15%) and a Communication Project (15%). The ESD booklet itself contained a detailed description for each work requirement to ensure that a school, while flexibly operating its own programme, was also working in terms of a set of nationally agreed constraints.

A Middle Ground on Assessment

The ESD team was concerned to ensure that a firm distinction be made between diagnostic, formative and summative assessment and to minimise the latter. For this reason, the number of common assessment tasks (five at year 12; four at year 13) were kept small, even though some of them were made up of more than one piece of work. While acknowledging the misgivings about standards-based assessment of educationists 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. This was in recognition of the value to student learning, especially in the process of formative assessment, of a statement of achievement objectives, and their place in guiding students to set their own language-related goals.

The design team, in proposing a reference test at year 12 and an external examination (Response to Text CAT) at year 13, acknowledged the need for an externally set and marked assessment task to moderate a significant proportion of a student’s work. At year 12, the reference test was proposed as a moderating device for 60% of the student’s work in total (i.e., the Writing Folio and Response to Text). At year 13, it was proposed that, in addition to the 30% of coursework examined and

moderated externally, a further 50% of coursework would be externally moderated. At year 13, only the Oral Communication CAT (20%) would be internally assessed and moderated.

The grade-related band descriptors (see Chart 1 for an example) 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 these work requirements were clearly linked to the grade-related band descriptors.

In a further attempt to reach the goal of acceptable consistency, this study design:

- built sets of constraints into the descriptions of the work requirements and CATs;
- designed generic marking guides (see Chart 1) for national use;
- made use of a “best fit” or “typicality” system to ensure that students are assessed for their grade trend across three or four assessment events for both Writing and Response to Text;
- developed exemplary rubrics, that is, interpretations of the generic marking guides which specify particular level indicators in relationship to particular language events, such as responding to a poem or writing an editorial (see Chart 2 for a short extract);
- made use of external moderation via sampling and check-marking;
- incorporated the use of exemplars as guides to the establishment of standards;
- emphasised the importance of professional networking among teachers as a way of standardising task-setting and grading.

Chart 1 is an example of a generic marking guide for Writing. It uses a ten-point grading scale to allow fine discriminations in terms of its 5-level grade descriptors. A student’s writing would be graded for two aspects (learning outcomes):

- the use of 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
- the development of confidence and competence in employing consciously the language features conventionally associated with a particular genre.

Chart 1 Marking guide for Writing

Grade	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 straight-forward 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 perceptive 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.

An extract from a rubric is given in Chart 2, which provides marking criteria for a particular written genre, 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 in order to gain a C or C+ in the “Conventions of Language” aspect of Writing. A similar set of indicators exists for the “Content and Context” aspect of Writing (not shown).

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.

Chart 2 Extract from rubric for the Expository Essay

Grade	Conventions of Language: (from 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.

The 1998 Year 12 English Study Design trial

By the beginning of 1998, the project team had had a widespread response from schools interested in trialing the ESD at both year 12 and year 13. Without the input of Government funding and given the reluctance of NZQA to bestow quasi-Bursary status on the year 13 ESD course, a decision was made to proceed with a trial of the ESD, funded by the University of Waikato’s School of Education Development Fund, at year 12 only, in a limited number of New Zealand secondary schools. The project also received financial support from Victoria University of Wellington (from the Research Committee of the Faculty of Humanities and Social Sciences) in order to conduct the evaluation of the ESD.

Thirteen schools eventually participated in the trial. Most of them simply redesigned their SFC English packages and, with NZQA’s accommodation, revised their SFC Course Statements to bring them in to line with ESD requirements (weightings, numbers of assignments and course componentry). A smaller number participated by incorporating a number of ESD work requirements into their sixth-form English programmes without adopting the ESD in its entirety.

For networking purposes, participating schools were arranged in three clusters, and initial cluster meetings were held in Invercargill, Christchurch and Auckland in February and March. By the time of the first round of cluster meetings, the project team had developed *A Practical Guide for Course Development and Assessment* (1998) which contained information on adapting a Sixth-Form Certificate English programme to the ESD, sample assessment tasks, and sets of marking guides and rubrics.

For the duration of the 1998 trial, Julie Moor, HOD English at Linwood High School, accepted the role of national moderator. Each school was asked to send her a range of assignments sampled from a number of common assessment tasks for check-marking and moderation purposes. Owing to funding restrictions, the project team decided against trialling the proposed reference test in 1998, even though most schools were keen on doing this.

Evaluating the 1998 ESD Trial

The evaluation was conducted towards the end of the school year and comprised three questionnaires, one for HODs, one for teachers of the programme and one for students.

Each questionnaire was developed during cluster meetings with schools in July 1998 and was based on themes deemed important for assessing whether the programme was working well. During cluster meetings, draft questionnaires were developed and then refined through comment from those present.

Towards the end of October, the complete set of questionnaires was sent to each school. Unfortunately, only ten of the thirteen participating schools were able to undertake the survey, and two of these omitted the student evaluation. One of the schools which did not supply data contacted the evaluator to apologise, identifying unforeseen factors which had limited the opportunity to collect the data (the temporary employment of replacement staff and students being involved in rehearsals for an end-of-year production). One of the two schools which

omitted the student evaluation component struck timetabling problems at the end of year, while the other did not sight the evaluation material until it was too late to administer the student component.

Results of the HOD questionnaire

The questionnaire to heads of departments asked only one question:

Your school has elected to adopt the English Study Design, or aspects of it, for its Sixth Form Certificate English programme in 1998. What are the key factors that contributed to your school's decision?

The results produced three main responses: the need to find a standards-based system of assessment which did not involve the use of unit standards; dislike for the grading system used for Sixth Form Certificate (even though ESD grades still required transformation to SFC grades at the end of the course); and the ease with which existing Sixth Form Certificate programmes could be adapted to the ESD. Other reasons given for adopting the ESD included its apparent manageability for teachers and students, the use of a simple external moderation system, coherence in programme content and structure, and a manageable assessment system.

Results of the teacher questionnaire

The teacher questionnaire focused on a number of themes which the cluster meetings deemed important for inclusion in the evaluation. The themes covered were:

- scope and coherence of the ESD
- workload and manageability
- flexibility, and
- assessment and moderation.

Each theme provided the opportunity for teachers to give a rating and provide comments. In all, 17 teachers completed the questionnaire, although in some cases items were omitted when the teacher was unsure of an answer or was not in a position to provide a response.

Most items were presented in the form of a Likert Scale (strongly agree to strongly disagree) and their results are presented in Table 1. The remaining items are summarised in the text. In respect of Table 1, the wording of each item is abbreviated to identify the theme of the question. Next to each item is shown the percentages of teachers giving positive responses ("strongly agree" + "agree"), neutral responses (a

mid-point rating) and negative responses ("strongly disagree" + "disagree"). For a more detailed breakdown of responses, readers are referred to the final project evaluation report (Hall, 1999).

Table 1 Statistical summary of items in the Teacher Questionnaire

Item	Rating (%)		
	Positive	Neutral	Negative
Scope and coherence:			
Meets the requirements of the national curriculum	100	0	0
Provides a coherent Year 12 programme in English	100	0	0
Workloads and manageability:			
Manageable for teachers	100	0	0
Manageable for students	94	0	6
Manageable for record-keeping	100	0	0
Flexibility:			
Allows variety in teaching and content	88	12	0
Allows for innovation in teaching and content	82	18	0
Allows students to show range of skills and knowledge	94	6	0
Assessment:			
Validity of summative tasks	100	0	0
Appropriateness of assessment weightings	69	19	12
Value of standards-based criteria	76	18	6
Grade scale provides scope for discrimination	88	6	6
Moderation:			
Moderation procedures are helpful	69	31	0
A moderation test would be useful	86	7	7

Scope and coherence of the ESD: Two items dealt with the scope of the ESD and its coherence with the national curriculum. The statistical results show 100% positive support for the ESD on both items. Comments focused on the adaptability of the programme, its structure and coherence, and its fit with current teaching practice.

Workloads and manageability: Three items focused on teachers' perception of workloads and manageability. All three items indicate a high level of satisfaction with the programme. The one school which reported a negative response for the manageability and workload of students was combining elements from two programmes – the workload for the students was thus higher than that for other schools. Comments generally reflected the positive reaction of staff as noted in the ratings.

Flexibility: Three items focused on flexibility. While all items indicate a high level of satisfaction with the ESD (there were no negative ratings), it is apparent from the comments provided by some teachers that not all components of the course were as flexible as hoped for, and that some aspects of the previous Sixth Form Certificate programme were seen as more flexible.

Assessment: Four items in Table 1 focus on assessment while a further two items (not shown in the table) sought information on: (i) the suitability of the grade scale for discriminating between students; and (ii) whether an alternative method of assessment (e.g., external examination, unit standards) would be preferred. The statistical information for this section indicated:

- a high level of satisfaction with the range of summative tasks covered by the ESD (positive responses 100%);
- less satisfaction with the component weightings (positive responses 69%; negative responses 12%);
- a belief that the standards-based criteria were mostly helpful (positive responses 76%; negative responses 6%) but not always able to be adapted to all tasks;
- general satisfaction with the 10-point scale for grading student work, although two teachers felt that the scale might be extended to 20 scale points, while two thought that 5 points were sufficient;
- strong support for the ESD as a system of assessment, although 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.

Comments on this section of the questionnaire were generally favourable and supported the statistical information. Some comments focused on the need to revise the assessment criteria (this has already been done for 1999) while two teachers reported dissatisfaction with the current weightings for components.

Moderation: Two items focused on moderation. It should be noted that three teachers were not familiar with the system of moderation used in the ESD and therefore did not answer this section. Results indicate that most teachers (69%) were satisfied with the ESD moderation procedures (the remaining teachers gave neutral ratings), but an even greater percentage of teachers (86%) indicated support for the introduction of a moderation test, although one teacher opposed this.

Additional comments by teachers: In addition to the questions shown in Table 1, teachers were asked to identify the three things they most liked about the ESD and the main changes they would like to see made to the programme. The positive factors most commonly identified about the ESD were:

- flexibility, adaptability and sensitivity;
- variety and diversity;
- manageability and workload;
- content and structure;
- marking schedules and criteria;
- moderation with other schools;
- the challenging nature of the programme;
- fit with the strands of the national English curriculum.

The changes that teachers most often suggested were:

- the need for revision/rewording of the assessment criteria;
- the need for further developments in moderation (e.g., introduction of an external test); and
- revision of the weightings for the components.

Results of the student questionnaire

The student questionnaire was developed with a view to students providing their teachers with feedback on how they had found the programme. The thinking behind the evaluation was that students have a perspective on their own learning which should aid schools in terms of their programme design and teaching. It is important to recognise that most students would not have experienced alternative year 12 programmes in English (e.g., unit standards) and therefore the results provide no comparative information. Strengths or weaknesses in the programme from the students' perspectives say nothing about the strengths and weaknesses of alternative programmes.

Participants at the July cluster meetings agreed that it would be useful to have a standardised questionnaire which all schools should use. This would enable schools to compare student reaction in their schools with student reaction overall, and thus suggest whether aspects of the programme might, or might not, require further tuning within their school. Because of the effort needed in analysing data, it was agreed that the questionnaire would comprise mainly "closed" format items. While this loses some of the richness that qualitative data

provide, given the large number of students involved, no practical alternative was available at this point.

The student questionnaire was structured to obtain general information about the programme overall and particular information about task components of the programme. Not all schools answered every question – where a component was not taught in a school the associated items were omitted.

Table 2 sets out the statistical information for each item, in an abbreviated form similar to that in Table 1. Most items followed the “strongly agree” to “strongly disagree” format. The table shows the percentages corresponding to positive, neutral and negative responses for each theme. Full statistical details are given in the final project evaluation report (Hall, 1999).

Table 2 Statistical summary of items in the Student Questionnaire

Item	Rating (%)		
	Positive	Neutral	Negative
General aspects of the programme:			
Quality of course organisation	53	31	16
Clarity of course information	54	31	15
Stimulation of interest	32	45	23
Manageability of workload		49	35
16			
Students' perception of their improvement in basic skills:			
Writing skills	51	33	16
Reading, viewing and listening	50	37	13
Assessment:			
Amount of assessment*	63	29	7
Diversity of assessment	59	32	9
Understanding of grade criteria	57	32	11
Understanding of grades obtained	54	28	18
Quality of comments and feedback	49	33	18
Value of ESD components for own development:			
Workfile for organisational skills	29	42	29
Workfile for judging own progress	27	40	33
Language Investigation Project for how language works	45	36	19
Communication Project for verbal skills	43	38	18
Communication Project for preparation skills	53	32	15
Communication Project for team skills	54	35	11

* 63% about right; 29% too much or too little; 7% far too much or far too little

It is very clear that the general level of satisfaction of students with the programme is less than that of teachers. This, of course, is hardly surprising and likely to be a finding for any programme as teachers are subject specialists and course designers while students are learners and recipients. It should also be noted that students are *required* to study Year 12 English – they do not have a choice as they do for other subjects.

As a rule-of-thumb, it is desirable that the percentage of positive responses for an item is at least three times that for negative responses. Although this may seem arbitrary, it provides a sufficient test of whether students by and large are satisfied with the aspect of the programme being evaluated. However, attention should also be paid to the percentage for “neutral” evaluations – if this figure is too large, more caution is needed in accepting a result which satisfies the above criterion.

From Table 2, it is clear that on balance students were satisfied with most of the general aspects of the programme (stimulation of interest being the exception), their improvement in the skills of writing, reading, viewing and listening, and the way their work was assessed in the programme (quality of feedback falling just below the criterion of satisfaction). However, it is very clear that the Workfile did not meet with students' approval in respect of aiding their organisational skills or improving their ability to judge their own progress. The results for these two items suggest that the ESD developers should give thought to how this component of the programme might be improved. It is also clear that the Language Investigation Project was not received very favourably, nor was the Communication Project convincing for improving the oral skills of students (as perceived by students).

Table 2 gives no information about the variation in satisfaction levels across schools. It is worth noting that some items produced a substantial variation. The most extreme case was for the first item in Table 2, namely, quality of course organisation. For this item, the satisfaction levels (% of positive ratings) ranged from 82% to 27% and the dissatisfaction levels (% of negative ratings) ranged from 0% to 49%. In respect of the value of the Workfile for developing organisational skills, satisfaction levels ranged from 63% to 18% and dissatisfaction levels ranged from 16% to 39%. This indicates that some schools were able to make this aspect of their programme work more effectively than others. The overall message from these results is that the sharing of ideas across schools might be a useful strategy to heighten the

effectiveness of the ESD. Professional development is an important component of any curriculum and teaching innovation.

Summary of the evaluation

In summarising the evaluation of the Year 12 ESD trial the writers wish to highlight in particular the following findings:

- confirmation from HODs that they were looking for, and found, a workable standards-based programme which did not involve unit standards and which could be easily adapted to their current sixth form programme;
- the manageability of the workloads for teachers and the general satisfaction of teachers with the ESD overall;
- the information provided by students on their experiences of the Workfile and the variation across schools in student reaction to the programme.

The latter finding suggests the need for the sharing of ideas across schools and the development of a system for disseminating good practice in the use of the different components of the ESD.

Implications for *Achievement 2001*

In terms of the broad picture it presented, *Achievement 2001* appeared to stake out a middle ground between those who had been advocating unit standards, or who were at least sympathetic to what unit standards were attempting to do, and those who either favoured national examinations or were downright critical of unit standards as a form of assessment suited to school curriculum subjects. *Achievement 2001* announced 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 are to be retained, generating credits at levels 1 and 3 respectively. However, 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.

The aspiration to take a “middle ground” is reflected in the suggestion that for each subject, half of the assessment will be by external examination or by centrally generated common assessment tasks. New “achievement standards” will replace unit standards in

“conventional” school subjects and students’ performance will be graded according to their success. Unit standards have been retained, but for “non-conventional” subjects only (such as Tourism), and these still earn credit towards the NCEA. A level 4 “Scholarships” assessment is mooted for highest achieving students. A student’s annual results will be reported to show credits obtained through both achievement standards and unit standards, grades from achievement standards and examination marks.

Achievement 2001 itself recognised that while the broad picture had been addressed, much work remained to be done. To this end, it announced the setting up of a Qualifications Development Group (QDG) – a germinal Board of Studies, perhaps – which would appoint subject specialist panels, to determine the exact nature of the proposed achievement standards and to make recommendations on the proportion of internal and external assessment for each subject. The QDG would also review the list of school curriculum subjects, determine the nature of the level 4 scholarship award, determine the best forms of national moderation, and decide on the nature of the level 2 common assessment tasks.

In our view, the credibility of the new qualifications will be crucially affected by the solutions arrived at with respect to the proposed achievement standards, future moderation procedures and common assessments tasks. It is with respect to these issues that we believe that the ESD project and the report on its 1998 Year 12 trial has a good deal to offer. In a number of ways, the ESD project has foreshadowed possible solutions to the questions left unanswered in *Achievement 2001*. Some of these are as follows:

1. *How might a balance of internal and external assessment be achieved?*

The ESD has already developed a formula for this which has gained acceptance among a significant number of English teachers. At year 13, for example, only 20 percent of the course is not externally moderated. Thirty percent of the course consists of an external examination. The remaining two Common Assessment Tasks (CATs) are written assignments worth 30 percent and 20 percent respectively. The ESD proposal is that these be internally marked and externally moderated. However, should the Ministry lay down a requirement that at least 50 percent be externally examined, then one of these common assessment tasks could be marked by an external panel of markers, raising the externally examined component to either 50 or 60 percent.

2. *How might a system of nationally prescribed common assessment tasks (CATs) be developed at Year 12?*

The English Study Design has not only provided a model for this. It has teachers using these common assessment tasks already with their students.

3. *What form might the new "Achievement Standards" take?*

This is a crucial question and the eventual credibility of the NCEA depends on the solution arrived at. In material released in November 1998 to a selected group of secondary principals (but not included in their official releases), the Ministry suggested that there be four achievement levels, reported as A, B and C (credit) and D (no credit). There is just a suggestion here of a desire to retain the pass-fail nature of unit standards, despite the introduction of merit and excellence. Given the large sums of money invested in the development of unit standards and their trialling, and the number of educationists who have invested time and reputation into this form of assessment, there must also be a temptation to "tinker" with already written unit standards as a way of developing achievement standards. In our view, an attempt to combine what is essentially a form of competence-based assessment with a system of grade-related, achievement-based assessment is like mixing oil with water. We would view such an attempt as counter-productive, as liable to compound the flaws associated with the attempt to write unit standards for subjects such as English, and as having the potential to provoke yet another backlash from disaffected teachers. It is also difficult to see how a system of "modified" unit standards will ease teachers' concerns about manageability.

Like *Achievement 2001*, the ESD is based around a commitment to assessment against intended outcomes. However, in avoiding the unit standards version of standards-based assessment, the ESD has taken on board the 1986 Department of Education report's suggestion of achievement-based assessment at five levels. Moreover, it has developed this form of assessment into a more rigorous and sensitive instrument using a 10-point assessment scale. In addition, the marking guides and rubrics developed as part of the ESD span both internally and externally assessed tasks. There is no longer an uneasy mix which uses one system for internal assessment and another system for examinations. Moreover, the marking guides can be translated easily into either grades or marks.

4. *What form might a national system of moderation take?*

As part of the 1998 Year 12 trial, the ESD team trialled a system of national moderation and check-marking for most common assessment tasks. This procedure was endorsed by participating teachers. As researchers we will be trialling and evaluating the proposed Year 12 reference test in 1999 in order to explore its usefulness as a moderating instrument in two CATs and to compare it with moderating via sampling.

5. *What about workload?*

An important touchstone of *Achievement 2001* is acceptable teacher workload. One of the key motivating factors in developing the ESD was to come up with ways of teaching English that would espouse standards-based assessment and yet would not undermine teacher morale and innovation with absurd workload pressures. The results of the evaluation of the ESD indicate that manageability is not a problem with participating teachers to date. This is an aspect of the ESD findings that the Ministry might wish to look at closely.

Conclusion

This paper has focused on the development, philosophy, evaluation and possible consequences of the Year 12 ESD trial. The ESD was developed to provide a standards-based programme in English which was independent of unit standards. The schools that opted into the programme identified with the need for a standards-based approach but rejected strongly the unit standards version on the grounds of philosophy, pedagogy and manageability. The schools also wanted a programme which would not require significant adaptation from what they currently offered. The ESD programme was seen to satisfy both of these broad concerns.

The evaluation of the trial identified a high level of teacher satisfaction with the programme. It could be said, of course, that the programme was "doomed to succeed", given the strong negative feelings of the participating teachers towards unit standards and their desire to adopt a standards-based approach which could be embedded in what they were already doing. This does not invalidate the results, however, but rather demonstrates the importance of taking teacher views into account when innovations are introduced. Virtually all aspects of the programme were evaluated favourably by teachers, with

satisfaction ratings ranging from 100% (the scope of the programme, coherence with the national curriculum, manageability for staff, administrative manageability and validity of assessment) to 69% (assessment weightings and value of current moderation procedures). The largest negative rating was 12% for assessment weightings, although the comments provided by a few teachers indicated that the assessment criteria might require simplification (this has already been actioned for the 1999 programme).

Not surprisingly, students rated the programme less favourably than teachers, although with the exception of items focusing on the Workfile (encouragement of organisational skills and value for monitoring own progress), positive ratings exceeded negative ratings usually by a factor of two or three to one. It is clear from the students' ratings that satisfaction with the programme varied considerably across schools. This suggests a need for schools to interact more regularly in order to share good practice in teaching and assessment.

In looking to the future, it is argued that the ESD has a number of features which the Ministry's Qualifications Development Group (QDG) might wish to investigate or monitor closely over the coming year. These include: the structure and composition of the ESD; the method of standards-based assessment it employs and the coherence of this with both internal and external assessment; the manageability of the teachers' workloads; and the moderation procedures employed, including the proposed trial of a reference test in 1999. The ESD design and evaluation team caution the QDG against taking the view that a simple modification of the unit standard system (e.g., by increasing the number of grade levels and making unit standards larger) is all that is needed in order to overcome the shortcomings of the existing system. The problems are more fundamental than this and require a completely new approach to standards-based assessment, including full recognition of its strengths and limitations.

References

- Churchman, R., & Hall, C. (1997). *Unit standards and professional education: A question of compatibility*. Victoria University of Wellington, UTDC Academic Development Series.
- Department of Education. (1986). *Learning and achieving: Second report of the Committee of Inquiry into curriculum, assessment and qualifications in forms 5 to 7*. Wellington: Learning Media.

- Elley, W. (1993). Curriculum reform: Forwards or backwards? *New Zealand Annual Review of Education*, 3, 37-49.
- Elley, W. (1995). Issues of objectivity in assessment: A plea for moderation. In Roger Peddie & Brian 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, September). Whatever happened to quality? *English in Aotearoa*, 30, 21-27.
- English Study Design Project. (1997). *English study design*. Hamilton: University of Waikato.
- English Study Design Project. (1998). *A practical guide for course development and assessment*. Hamilton: University of Waikato.
- Hall, C. (1994). Obstacles to the integration of university courses into the National Qualifications Framework. *Higher Education in New Zealand, Occasional Paper Number 1*. Wellington: Syndicate of Educational Development Centres of New Zealand Universities.
- Hall, C. (1998). The national qualifications framework Green Paper: What future for the framework? *New Zealand Annual Review of Education*, 7, 29-57.
- Irwin, M. (1994). *Curriculum, assessment and qualifications: An evaluation of current reforms*. Wellington: Education Forum.
- Irwin, M. (1997, August). *The National Qualifications Framework: Where to now?* Paper presented to the Waikato Forum on Education, University of Waikato, Hamilton.
- Lennox, B. (1995, December). *Advocacy, evolution and learning: School assessment for the National Qualifications Framework*. Paper presented at the NZARE Conference, Palmerston North.
- Locke, T. (1997). An essay on assessment: The English casebook. *Waikato Journal of Education* 3, 93-117.
- Roberts, P. (1997). A critique of the NZQA policy reforms. In Mark Olssen & Kay Morris Matthews, *Educational policy in New Zealand: The 1990s and beyond* (pp. 162-189). Palmerston North: Dunmore Press.
- Tuck, B. (1995). What is wrong with standards-based assessment? In Roger Peddie & Brian Tuck (Eds.), *Setting the standards: The assessment of competence in national qualifications* (pp. 59-78). Palmerston North: Dunmore Press.

The authors

Cedric Hall is Professor and Head of School, School of Education at Victoria University of Wellington. He was formerly Director of the University Teaching Development Centre at the same university. His research interests include course design, teaching, learning, assessment and evaluation. Some of his recent papers have included a focus on quality assurance in higher education and developments in the National Qualifications Framework.

Terry Locke is Senior Lecturer in English Language Education at the School of Education, University of Waikato. His research interests include the history and rationalisation of English as a subject, the teaching of argument and critical literacies. His publications include a range of textbooks on aspects of the teaching/learning of English at secondary level.