Marking Writing Tasks for Tests of Literacy in Years 3, 5 and 7: Questions and Challenges

Lee Willett

Queensland
Studies Authority

Introduction

Writing is a fundamental aspect of literacy. It seems almost superfluous to say that we want to know how well students do it. While we want to know how well students can control the surface features of spelling and punctuation, we also want to know how well they control the structural features of generic structure and clause structure as well as how they label and order ideas. The writing that really shines demonstrates how well the students understand and respond to the social factors that shape writing - the demands of the task and the needs of the audience. But if students are to function well in society, if Australia is really to develop as a knowledge economy then students must understand that writing is also about the construction of knowledge. Through writing, we make patterns of thinking concrete and knowledge communicable.

For teachers and curriculum developers understanding how students acquire and develop control over the different aspects of writing is critical. The challenge for the testing program is how to conceptualise these dimensions in ways that are measurable, manageable and reportable. The challenge too is to measure writing in ways that support growth in teacher knowledge of writing assessment and which have the best possible curriculum effect. The changes we have made to the writing task are part of our efforts to show teachers how our testing program can help them develop powerful insights into developmental learning.

Background

This paper focuses on the writing task component of the Queensland Tests of Aspects of Literacy in Years 3, 5, and 7. The marking rubric used to mark the writing task evolved from the early days of the testing program when Queensland had a population Year 5 test and a sample Year 3 test. It was developed to measure and report on the "language in use" model underpinning the *Years 1-10 English Syllabus*. The context-text model was highly influenced by Systemic Functional Linguistics, which sees texts as socio-cultural constructions (Halliday, 1994). In a narrow interpretation of that model, the writing task was very much genre specific. Student writing in response to a genre-specific task was marked according to understanding of two contextual features - subject matter and awareness of the roles and relationships between the writer and the audience - and, six textual elements - generic structure, grammar, cohesion, vocabulary, punctuation and spelling.

Initially, the criteria were developed across six standards with a seventh added as Year 7 was added to the test program. This created the potential for a large number of mark points. In addition, as the benchmark descriptions of the writing task became available descriptions of particular levels were altered to reflect the substance of the benchmark statements. An analytical approach to the marking of writing underpinned the use of the rubric.

Experience, the inclusion of the Year 7 students, and the shift from sample to population testing of the Year 3 students had a big impact on the testing program. The lessons of hard-won experience and reflection caused us to question the value of using an analytical approach for a single piece of first draft writing created in 45 minutes. How valid was it to make such fine-grained analytical decisions on this basis? The fact that some of the dimensions mapped by the marking grid were dependent variables with resultant high correlations was also an issue. Analytical marking is time consuming and the marking rate associated with it is costly. Therefore improving the marking efficiency through combining dimensions that measure similar characteristics can be financially and logistically beneficial.

Other problems arose from using a single marking rubric for marking all year levels. This was a manifestation of the need to locate student performance across all year levels on a common scale. As practising teachers, the markers have a sense of a 'good piece of writing' or a 'poor piece of writing' relative to a particular year level. A valuable classroom skill, this sense was a major source of discrimination problems when, despite criteria being laid down, some markers awarded Year 3 and Year 5 students marks from the highest levels of the grid – a level that represented a very high, polished standard of writing, a grade the scripts did not warrant.

For both measurement and educational reasons, the decision was taken to move to impressionistic marking. However, we faced some difficulties, not the least of which was that both the markers and teachers liked the existing analytical rubric. Teachers felt that it gave them good information about growth in students' writing and that they could apply this knowledge to classroom assessment tasks. Markers had internalised the marking grid and felt comfortable with it.

The fact remained, that by 2003, we were using a marking rubric that had become somewhat patched. While the number of marking criteria had been reduced to three, contextual features, textual features and spelling, teachers also made judgments against the other criteria that had originally contributed to the mark so marking was time consuming. This reduction in criteria also meant that there were not enough mark points to discriminate adequately between student performances over the four years of schooling covered by the testing program. However, while collecting data and marking on three criteria, the writing rubric still retained its analytical format with all of the sub-strands the markers used to make their decisions.

For 2004, the decision was made to change the approach to the design of the marking rubric. The change was to be incremental, using each change to evaluate whether the changes were being made in the right direction.

Rationale for a new marking rubric

A writing rubric needs to produce reliable and valid data that discriminates well between different levels of student performance and must be quick and straightforward to use. This is easier said than done.

In developing a new marking scheme, two approaches were open to us. One involved a major change: to mark holistically against a small number of criteria such as was done in the IEA study of written composition (Gorman, Purves & Degenhart, 1988) or is done in the 6+1 trait model developed by the North Western Regional Educational laboratory and now used in the testing regimes of some American states. These scales have merit in their ease of use and in the way they relate to common-sense notions of writing held by the community and much of the profession. They have the advantage of setting an articulated standard. Where the markers share their understanding of a standard this adds to the consistency of marking. Experience with the Queensland Core Skills (QCS) Test undertaken by Year 12 students supports this statement. However, the use of teacher judgment and moderation in senior secondary assessment has meant that the markers come to the writing task of the QCS Tests with a shared knowledge of the standards and the underpinning assessment practices. This is not the case with primary teachers whose assessment has been largely formative and, at best, school-based.

Another issue with the use of the first approach is that, like much of the way we learn, learning to write involves a certain amount of revisting and redefining of knowledge. As students shift their learning focus to a new learning problem earlier demonstrations of related learning may appear to disappear. This is often talked about as redundancy in learning. One clear demonstration of this is students' control of punctuation. Years 2 and 3 students writing simple sentences may well show control over sentence boundary markers, but as they attempt to build clause-complexes, students frequently lose

their sense of where the clause boundaries are. This returns when students develop their understanding of how clauses relate to each other and the role punctuation has in marking out the semantic boundaries in texts. It is counter-productive to punish its temporary absence of learning which is in fact a signal of students' engagement with the deep structures of writing and a sign of positive growth.

Similarly, there appears to be a revisiting of some aspects of clause structure as students develop their control over it. In an examination of the 2002 writing task, we were somewhat surprised by the low level of use of subordinate clauses in the Year 5 scripts. We had expected to see more frequent and better examples than those produced by Year 3. By and large, the use of subordinate clauses was still quite basic, but students seemed to be writing simple sentences with increased lexical density using phrases and some 'packed' noun groups. It must be said that this was a preliminary scan of the scripts and students' writing may have been the result of a cohort or task effect. However, it may be that this is the beginning of the journey writers make as they learn to pack the lexical density within clauses. It does indicate a probable need to plan some level of redundancy into the teaching of clause structure in writing.

Perhaps the most difficult challenge in conceptualising the framework of a marking rubric for writing is that there is no end point. As expertise develops, old challenges are met and new challenges in response to new complexities arise.

If the writing task is to be used to improve our understanding of how students develop as writers and to map their progress, then the key markers of writing development need to be identified. For this to happen, more than a competency-based approach to the assessment of writing is needed. For the criteria used to mark the writing task to be useful to classroom teachers, they need to map the territory, describing not only the qualities of a particular script but also providing an indication of the next step on the learning continuum.

Development of the writing rubric

The writing rubric was developed within the practical constraints that always exist. These ranged from mundane budget issues to measurement and operational issues. The development of a new, outcomes-based English syllabus also influenced the construction of the rubric. In writing the syllabus outcomes, the syllabus writers needed to identify and describe the major indicators of writing development and to attribute a level to them. The confluence of testing and syllabus development objectives has provided a great opportunity to test and refine ideas about the development of writing and so develop a closer match between the syllabus outcomes and the testing program.

As a result, the decision was made to develop the rubric to match the anticipated levels of the new English syllabus. In response to the need to have an increased number of mark points and to deal with the difference in time between when the tests are conducted and when the level outcomes are to be achieved, an intervening level was developed. Thus descriptions were developed for the first five levels of syllabus outcomes, together with five intervening levels.

To deal with scripts which have no response or an unmarkable response, two additional levels were added. The approach to selecting and ordering the rubric content was framed by the adoption of a writing task that was not genre-specific but allowed students to make the decisions about the tasks and to respond in a small range of appropriate ways. The wording and the sequencing of the criteria were constructed to account for this.

The rubric is constructed for impressionistic marking against a standard. Four major threads are woven through the criteria describing the contextual factors and textual features. These are

- understanding of the social and cultural purposes of writing
- understanding of audience
- overall structure
- clause structure.

These are the basis for the development of rubrics for use in marking any writing task. Other textual features such as cohesion, vocabulary and punctuation are referred to in the descriptions. Because spelling in writing is combined with dictation and proofreading to produce a spelling score, spelling is the third criterion according to which the writing is marked. The spelling descriptions are framed by the stages described by Templeton and Bear (1992, 2002). The descriptors for this year's spelling criterion are unchanged from those used last year. The spelling data presented in Tables 3, 6 and 9 are indicative of discrimination issues associated with the old writing rubric.

Descriptions of the first two threads were used to construct the criterion describing contextual factors. Students' understanding of purpose is described along a continuum beginning with little knowledge of purpose, through to knowing writing is for communication and finally to knowing that writing is a tool for influencing people and constructing knowledge. Bereiter and Scardamalia (1987:6) describe novice writers as moving from conversation, to knowledge telling and then to knowledge transforming. Similarly novice writers' understanding of the demands of audience can be described as a continuum, identifying the role students take and their development of audience awareness as they shift from self-audience, to trusted and known audiences, and finally to identifying and using the knowledge and values of audiences to frame and achieve the purposes for their own writing (Moffatt, 1981).

One of the critical milestones in the development of awareness of audience in novice writers is the understanding of the difference between written and spoken language. In their study of novice and expert writers, Bereiter and Scardamalia (1987:90) assert that novice writers need to transcend their dependence on a conversational partner. In their early writing, novices produce writing that can be read as conversational exchange. These conversational exchanges are particularly obvious to markers. The conversational nature of early writing is particularly evident in the development of clause structure. For example,

> *I learnt to look after some mice and then they had babys and the babies grew up and escaped and I had to catch them.* (John, aged 12)

Novices need to learn the difference between the ways propositions are linked and ordered in written text. The holes and superfluous details in the subject matter are also a manifestation of this lack of knowledge about writing and the need for a 'conversational partner'.

Because the development of clause structure is such a clear marker of this conversational style and is something novice writers do in all forms of writing, it was used as a major thread in the description of the textual features. The construction of the marking rubric was influenced by Gunther Kress's descriptions of the development of clause structure in students' writing. Kress (1994) suggested stages of

- pre-conjunction (sequence alone)
- rudimentary (conjunction plus sequence)
- subordination
- embedded clauses (hierarchical and logical order predominant).

In marking scripts, markers find the early aspects of mastery over clause structure fairly easy to identify as illustrated by the example in Figure 1 where the sequence alone is used to imply the relationships

between the propositions in the text.

Dear Misd HALStR

get Rid of the yomfs ples I hat even gomfs in the school ples get rid of the. even day wen school fishis I get home It tas hisn of lottime to get Into noml. your In slely sophic

Figure 1: Pre-conjunction

As mentioned, in 2003, we undertook a small review of the scripts from the previous year, comparing the attributes of scripts of students with particular scale scores. As well as the apparently small amount of growth of use and control over subordinate clauses, there was an appreciable difference in the nature and quality of simple sentences. Students use more phrases and noun groups to increase lexical density. This may indicate the beginning of students' ability to increase the lexical density of clauses. As such, it may represent a stage necessary to developing the clause complexes evident in more sophisticated writing (Figure 2). This was added to the descriptions (standard E) in the rubric.

> Over the last term,
> I have learnt all the ways and
> doings of the college, but one issue has
> not yet failed to puzzle me; School
> jumpers.
> I understand that the regulations
> regarding uniform are set and are to
> be followed, but the rule that states,
> that jumpers cannot be worn to and
> from the grounds unless accompanied
> with a blazer is absurd! Basicly, if
> we just want a bit of extra warmth, we
> have to wear just the blazer!
> Don't get me wrong, the blazers
> are nice but if I were to wear the blazer
> to school every cold morning, I may as
> well super-glue it to me and be done
> with! Unlike jumpers, you can't just

Figure 2: Embedded clauses used to increase the lexical density

Another major thread used to link the descriptive criteria is that of overall text structure or organisation. Previously this thread was developed in terms of control of the selected generic structure. However, at a surface level, students understand the general organisational shape of the text. It is the nuances of the structure and an intervening level of the macro-proposition, which is played out in the tightness of the organisation and the lexical cohesion that is more problematic. Students showing more control of this intervening stage of writing are able to develop their writing tightly about a single theme (Figure 2). This ability to develop and tightly link the macro-propositions through the text, to render in linear form ideas which may not be linear is a mark of writer moving toward maturity. This requires a major transformation both in the way knowledge is structured in text and of the knowledge students need if they are to manage this step in the development of their knowledge of writing.

These identified markers of writing were used to construct a continuous scale for potential use in marking writing across Years 3, 5 and 7. The scale was divided across the three year levels, with two standards appearing in common with all of the year levels. The rubrics for each of the year levels appear in Appendix 1.

Using the rubric

The increased number of marking points required some changes to the presentation and use of the rubric. The markers were given a restricted set of six standards to mark against. Each of the standards was colour coded with the two standards describing the syllabus outcome statements in two shades of the same colour. This was done to assist markers to internalise the standards and navigate the rubric. After extended discussion between item writers and measurement officers, the decision was taken not to truncate the scale but to allow markers to move up or down as they thought necessary with the caveat that any scripts marked outside the restricted range be subject to double marking by the senior markers. This was an added quality control measure.

Markers were given approximately six hours training in the use of the rubric. This involved three sessions:

- an introduction to the rubric, its philosophical basis, the changes and the approach underpinning the rubric.
- a whole group activity – introducing a set of annotated scripts.
- application of the rubric in a small group setting.

This was followed by the collection of baseline data on individual markers.

Markers were asked to use the rubric to make a 'most-like' judgment. If necessary, annotated scripts were provided to markers to reinforce the articulated standards.

Because one of the purposes of the marking operations was to test how well the rubric could reflect the hypothesised levels of the developing English syllabus, markers were not told of the link between the levels in the marking rubric and the learning outcomes of the draft English Syllabus.

Results

As a measurement device, the rubric appears to have worked reasonably well. The marking operation went quite smoothly and quickly. The rubric appears to have been adequate in discriminating between different levels of student performance. There is a positive progression across the scale of the mean ability levels of the students and the thresholds. That this might have occurred through 'marker centring', or through markers simply rating 'good students' higher regardless of the match to criteria, does not appear to be borne out by the data or by the explorations we have done in re-evaluating scripts at some of the higher levels.

Markers felt that the rubric was easier to use than the previous rubric and allowed for better descriptions of students' writing. They felt the rubric enabled them to make more accurate judgments about the scripts and to work more quickly. They certainly liked the use of colour to mark out the standards and in discussing scripts, conversational comments such as "I'm thinking yellow here" could often be heard.

The discussion of results is framed against the evaluation of how the syllabus outcomes, as articulated by the marking rubric, performed. Evaluation of the writing performance against the level statements needs to be considered against the fact that this is a single piece of first draft writing produced under the stress and restrictions of the test situation. This evaluation has answered some questions and raised new ones.

The results for Year 3 students suggest that the outcome statements for this level have, in this exercise at least, been validated. Because the outcome statements represent a broad band of performance, the outcome statements were divided into a higher level, standard D, and lower level, standard C. The data for contextual factors (Table 1) show that, on the criteria for contextual understanding, 40 per cent of the Year 3 students produced writing rated standard D or above while 79 per cent produced writing standard C or above.

Table 1: Year 3 Writing—contextual factors

| | | | | | Infit MNSQ | 0.88 | |
| | | | | | Disc | 0.67 | |
Categories	A	B	C	D	E	F	G
Count	2149	8256	19676	15512	3985	580	38
Percent	4.3	16.4	39.2	30.9	7.9	1.2	0.1
Mean Ability	-1.92	-1.23	-0.50	0.19	0.84	1.48	2.04
StDev Ability	2.46	0.68	0.74	0.76	0.80	0.86	1.15

Similarly, the data for textual features (Table 2) show that 35 per cent of the Year 3 students produced writing at standard D or above, while 77 per cent demonstrated the level of control over textual features expected of Year 3 students by the syllabus writers.

Table 2: Year 3 Writing—textual features

| | | | | | | Infit MNSQ | 0.87 |
| | | | | | | Disc | 0.68 |
Categories	A	B	C	D	E	F	G
Count	1907	9937	20888	13700	3306	438	16
Percent	3.8	19.8	41.6	27.3	6.6	0.9	0.0
Mean Ability	-2.04	-1.14	-0.45	0.28	0.96	1.53	2.55
StDev Ability	2.57	0.71	0.75	0.75	0.79	0.90	0.66

The spelling criterion was unchanged from the marking rubric used for the previous year's writing task. The results are included (Table 3) because they are part of the writing task and so need to be reported. They demonstrate that if useable data are to be produced to inform teaching and learning, the number of standards must be increased. This is the focus of current work by a team of syllabus writers, spelling experts and test developers who are engaged in conceptualising low, middle, and high bands of the Bear and Templeton Stages. This is something the authors themselves have done (Bear et al, 2002) and this approach has the potential to account for some of the redundancies in the way students learn to spell. It also allows for an integrated description of the spelling strand of the literacy test.

The data show that the vast majority of students are grappling with mastery over the patterns described by Bear and Templeton's letter naming and *"with-in" word* pattern stages.

Table 3: Year 3 Writing—spelling

| | | | | | | Infit MNSQ | 0.72 |
| | | | | | | Disc | 0.71 |
Categories	A	B	C	D	E	F	G
Count	322	5 689	21 017	21 308	1 819	40	0
Percent	0.6	11.3	41.9	42.4	3.6	0.6	0.0
Mean Ability	-2.66	-1.62	-0.72	-0.28	1.29	2.03	0.00
StDev Ability	6.06	0.59	0.69	0.75	0.79	1.08	0.00

Compared with that of Year 3 students, there is a marked difference in the level and range of writing performance of the Year 5 and Year 7 students. For Year 5 students, the markers were given standards C to H as the basic marking range. The standards E (low Level 3) and F (high Level 3) represent the syllabus outcomes expected by the end of Year 5. While the performance is more evenly distributed than in Year 3 (Table 4), the number of students performing at or above the expected level of outcomes is lower—51 per cent are at or above the expected level for Year 5, in their understanding of the contextual factors that influence writers and writing.

Table 4: Year 5 Writing—contextual factors

| | | | | | | | | | | Infit MNSQ | 0.92 |
| | | | | | | | | | | Disc | 0.71 |

Categories	N	A	B	C	D	E	F	G	H	I
Count	168	507	1 642	7082	16 590	14 733	8 405	3 005	434	49
Percent	0.3	1.0	3.1	13.5	31.5	28.0	16.0	5.7	0.8	0.1
Mean Ability	-2.55	-1.98	-1.69	-1.01	-0.40	0.14	0.64	1.20	1.78	2.50
StDev Ability	0.90	0.79	0.60	0.68	0.69	0.70	0.71	0.76	0.82	1.01

The data suggest that similar levels of performance to this are represented in the level of control students have over their deployment of the textual features of writing (Table 5). Here 48 per cent of Year 5 students produced writing at or above the standards representing the outcome statements. This is significantly less than the proportion of Year 3 students who achieve their expected levels. However, for both criteria, contextual factors and textual features, there are another 31per cent who are close to achieving the expected level.

Table 5: Year 5 Writing—textual features

| | | | | | | | | | | Infit MNSQ | 0.89 |
| | | | | | | | | | | Disc | 0.71 |

Categories	N	A	B	C	D	E	F	G	H	I
Count	168	417	2 106	8 244	16 634	15 508	7 065	2 030	414	29
Percent	0.3	0.8	4.0	15.7	31.6	29.5	13.4	3.9	0.8	0.1
Mean Ability	-2.55	-2.21	-1.56	-0.96	-0.36	0.23	0.75	1.36	1.76	2.92
StDev Ability	0.90	0.60	0.66	0.69	0.68	0.69	0.71	0.75	0.82	0.86

The spelling data (Table 6) suggest that the vast majority of Year 5 students are still developing their understanding of the *"with-in" word stage* patterns as well as the syllable juncture. This includes such elements as long vowel patterns and complex consonant patterns. However, as previously stated, the measure is not sufficiently discriminatory to be used to identify the ways in which this performance is qualitatively different from that of the Year 3 students. The results of the dictation and proofreading items show that such a qualitative difference does exist. While Year 3 students may have reasonable control over these patterns in single syllable words, the added complexity of understanding and deploying these letter patterns in multi-syllable words means that Year 5 students are still refining their understanding of the "*with-in word*" patterns although this happens in a qualitatively different and measurable way.

Table 6: Year 5 Writing—spelling

| | | | | | | Infit MNSQ | 0.71 |
| | | | | | | Disc | 0.71 |
Categories	A	B	C	D	E	F	G
Count	205	1108	7 567	30 443	12 505	742	41
Percent	0.4	2.2	14.4	57.9	23.8	1.4	0.1
Mean Ability	-2.52	-1.98	-1.20	-0.18	0.79	1.71	2.20
StDev Ability	0.89	0.53	0.61	0.71	0.72	0.83	0.88

In marking the Year 7 scripts, the markers were given the standards E to J with the outcomes for syllabus Level 4 represented by standards G and H. However, the markers have made use of all levels of the grid. Because all scripts below standard E have been double marked, it is probable that the results accurately represent the distribution of performance.

The trend evident in the Year 5 data continues with an increase in the spread of performance (Table 7). Here 43 per cent of students produce writing of an adequate standard to meet those described aspects of the syllabus outcomes concerned with contextual understandings. However, the 26 per cent of students achieving standard F may well be able to meet the expectations of the outcome standards under classroom conditions.

Table 7: Year 7 Writing—contextual factors

| | | | | | | | | Infit MNSQ | 0.92 | |
| | | | | | | | | Disc | 0.71 | |
Categories	N	A	B	C	D	E	F	G	H	I	J
Count	52	159	356	1266	4395	10366	14054	13881	6404	2177	119
Percent	0.1	0.3	0.7	2.4	8.3	19.5	26.	26.1	12.0	4.1	0.2
Mean Ability	-2.62	-1.81	-1.96	-1.53	-1.00	-0.47	-0.03	0.44	0.90	1.46	2.02
StDev Ability	0.75	1.03	0.56	0.58	0.62	0.64	0.65	0.65	0.67	0.79	0.73

In Year 7, the data for textual features (Table 8) indicate a wider gap between students' ability to show evidence of their understanding of the contextual features that shape writing and their ability to control these textual features themselves. Some 36.0 per cent of the students are at or surpass the level of performance described by standards G and above. It should be remembered that another 26 per cent of students in standard F are working close to that level. While the reason for the emerging gap between the two might be attributable to the number and sophistication of different textual features expected from writers, the gap may well be due to an increasing difference in the maturation levels of students.

Table 8: Year 7 Writing—textual factors

| | | | | | | | | | | | Infit MNSQ | 0.90 |
| | | | | | | | | | | | Disc | 0.72 |

Categories	N	A	B	C	D	E	F	G	H	I	J
Count	52	103	477	1559	4595	13 261	14 014	11880	5615	1411	63
Percent	0.1	0.2	0.9	2.9	8.6	24.9	26.3	22.3	10.9	2.7	0.1
Mean Ability	-2.62	-2.42	-1.77	-1.49	-0.95	-0.39	0.08	0.54	0.97	1.66	2.17
StDev Ability	0.75	0.49	0.67	0.58	0.62	0.65	0.64	0.64	0.68	0.77	0.74

On the face of it, the spelling data (Table 9) show a marked similarity in performance to that in Year 5 and the same caveats about the quantitative difference in performance apply. What is striking here is the number of students (8%) demonstrating control of the letter patterns associated with Bear & Henderson's *Derivational Constancy Stage*. Understanding of and control over the patterns of this stage is the last great hurdle before mature and correct spelling. This result is consistent with what the research says about older students' control over English orthography (Templeton, 1992; Goldsmith, 1986). This has implications for the way we teach students about the spelling system in the upper primary school and particularly what we do about the continuing teaching of spelling in the secondary school.
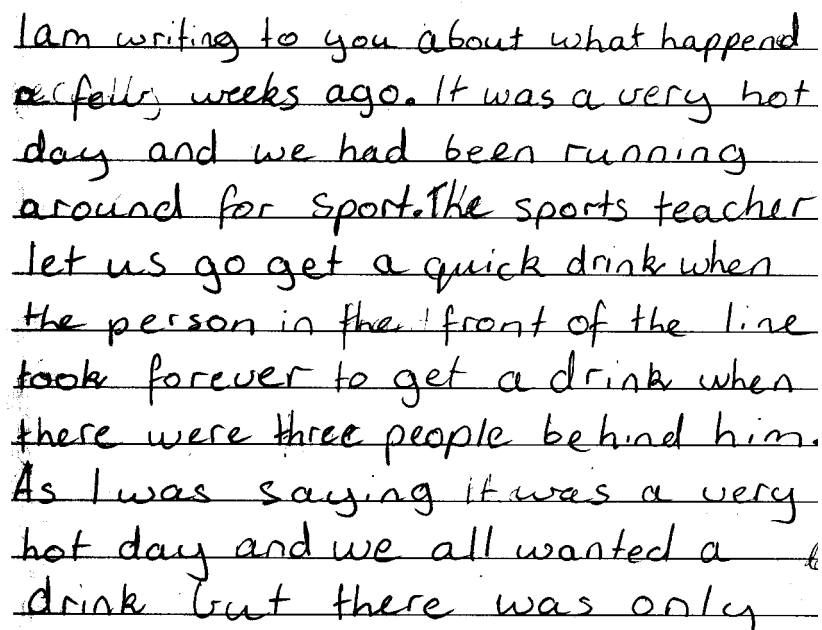
Table 9: Year 7 Writing—spelling

| | | | | | | Infit MNSQ | 0.71 |
| | | | | | | Disc | 0.71 |

Categories	A	B	C	D	E	F	G
Count	73	220	1961	20 166	26 393	4175	241
Percent	0.1	0.4	3.7	37.9	49.6	7.8	0.5
Mean Ability	-2.62	-2.21	-1.48	-0.49	0.39	1.24	1.97
StDev Ability	0.67	0.46	0.59	0.69	0.70	0.72	0.89

Discussion

There is a qualitative difference in the performance of the Year 3 students on the one hand and the Years 5 and 7 students on the other where declining numbers of Years 5 and 7 students met the anticipated outcomes. This raises a number of questions. However, it is important to remember that the writing being measured against the rubric standards is one piece of first draft writing, conducted under pressure, at a point in advance of the time at which these students are expected to demonstrate the desired learning outcomes. Because of this, some lag in the expected performance against the outcome statements was anticipated. That this lag in performance is more marked in Year 5 and in Year 7 than in Year 3 needs to be considered on a number of levels.

The first and most obvious consideration is that the standards at the top end of the scale may need to be redefined. Certainly, this is an issue we will explore through the sample selected for further qualitative evaluation. However, the statistics are reassuring and suggest that only minor adjustments will be necessary and that there may be other questions to consider in an evaluation of the rubric.

The distribution of the data raises the questions of whether the identified milestones of development are the right ones and placed appropriately in the sequence. One of these milestones was the point at which students demonstrated control over basic subordinate clauses. On the rubric, this milestone was described in the shift between standards E and F. The writing described in standard E had simpler more conversational structures while that at standard F had more developed clause structure moving toward writing that was more 'knowledge telling'. However, it may be that markers have difficulty in discerning the shift between these levels. This may be due to the fact that vestiges of conversational writing in facets other than clause structure remain in students' writing. The writing described by standard F shows that some of the attributes of 'conversational writing' are present while at the same time demonstrating evidence that the students are developing control over the subordinated clause structures of written language. In the example shown in Figure 3, there is some evidence of the writer's need to fill in information for the reader and of some mastery over basic subordinated clauses.
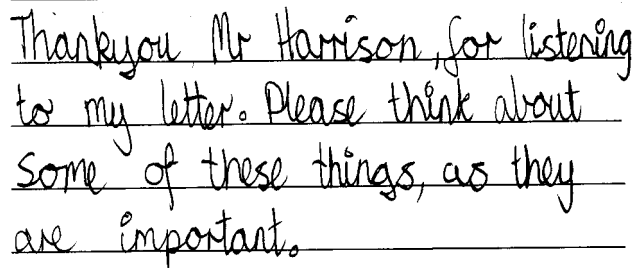
I am writing to you about what happened a (few) weeks ago. It was a very hot day and we had been running around for sport. The sports teacher let us go get a quick drink when the person in the front of the line took forever to get a drink when there were three people behind him. As I was saying it was a very hot day and we all wanted a drink but there was only
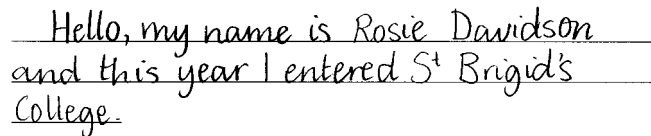
Figure 3: Developing subordination

However, the tenor of the language, the ordering of the subject matter and the conversational move signalled by the lexical choice '*as I was saying*' suggests that although this student is developing awareness of audience, he has not yet moved totally beyond writing for a conversational partner. This

was a common feature of writing at this stage of development. Even well-crafted pieces of writing still had single instances of lexical choice, which still signalled the writer's developing awareness of audience and written text.

> Thankyou Mr Harrison, for listening to my letter. Please think about some of these things, as they are important.

> Hello, my name is Rosie Davidson and this year I entered St Brigid's College.

Figure 4: Lexical choices suggest continuing conceptualisation of 'the conversational partner'.

This may be affecting the way markers assign grades and this is an issue to be investigated when the sample of this year's tests is drawn. There may be a need to differentiate more precisely the descriptions of standards E and F. The interesting issue for writing pedagogy and curriculum development is the possibility that students' need for a conversational partner continues for much longer than anticipated.

There are other possible explanations for the differences between the writing performance of the Year 3 students and those in Years 5 and 7. In their study of writing from the Maryland testing program, Goldberg, Roswell and Michaels (1996) evaluated the effectiveness of including in a testing program those writing processes that were common classroom practices. In Maryland, the writing task included a revision stage in which students were allowed to use strategies such as peer conferencing to produce a final draft of their writing. The evaluation of Goldberg et al (1996) was that in general it made very little difference. In the case of Year 3 students, they found that when opportunities for revision and editing mirroring classroom instruction were introduced into the testing regime, the final drafts of Year 3 students were 'more likely to give a less accurate picture of students' writing abilities than do the rough drafts'. Furthermore, they found that the writing samples that Year 3 students produced under test conditions were more likely to be an accurate representation of their writing ability.

At the same time, Goldberg et al (1996) found that older students engaged in very formulaic editing practices that led them to make only surface level changes that often made very little difference. More worryingly, they found that older students were unable to provide feedback about the rhetorical devices that would aid revision and quite often removed those that gave shape to the text. This finding could well be one possible explanation for the difference between the Year 3 results and those in Years 5 and 7. Further qualitative explorations around this issue need to be undertaken to determine whether this is the case and to determine the pedagogical implications.

Goldberg et al (1996) also noted that in revision, older writers were more likely to pay increased attention to the generic structure and to purpose, but were unable to engage with major reordering or rewriting of sections to make the meaning clearer. This was a developmental issue raised by Bereiter and Scardamalia (1987) when they described the difference in the way novice and expert writers recognised the relevance and weighting of ideas and were able to use this knowledge to shape their writing. This was particularly pertinent to the decisions writers make as they reproduce nonlinear ideas as linear text. Doing this requires a major transformation in the way knowledge is structured in text. That this is an issue which needs to be considered in the progression between Standards G and I, is evident in the difference in length of scripts. The best scripts matching the rubric standards I and J are shorter and more succinct. Those in the middle range are longer and tend to have problems of coherence.

The last of the considerations offering an insight into the differential performance between the Year 3 students and those in Years 5 and 7 also is suggested by the Bereiter and Scardamalia study (1987). They describe students in the middle years of schooling as taking one of two paths as writers. One group of students take what Bereiter and Scardamalia (1987) called the high road. These writers actively develop the cognitive, problem-solving tools that allow them to progress to more complex levels as writers. The other group take the low road, concentrating on the surface features that allow them to produce a neat error-free piece of writing. The difficulty with this is that having done so, they still need to grapple with the complex problems that confront mature writers and appear to be hesitant to do so. The relative change in performance between Year 3 and Year 5 and 7 raises the question whether we have numbers of students who have taken the low road and whether, as a result we need to examine writing pedagogy to make sure that these students do engage with the more complex aspects of writing and develop the knowledge and associated problem solving ability to develop as writers.

Marker considerations

Feedback from the markers, which was supported by the data, was that the markers found the A to E and the H to J range of the rubric easy to mark. In addition to those already identified, there could be reasons associated with the markers that could account for this. Writing of the kind defined at each end of the scale is easy to identify and describe. Experience from working with teachers with other assessment regimes such as the *Year 2 Diagnostic Net* and with markers in previous marking operations indicates that teachers find the progress students make through the conversational stage of writing easy to identify. At the other end of the scale good writing is also easier to identify. However, discriminating between levels of competence that fall between these extremes is more difficult. This requires a more detailed knowledge of grammar, cohesion and structure. Next year, we will take these aspects into account in redefining the rubric and the annotated scripts and as a focus for marker training. In addition, we are considering the development of marker recruitment training where potential markers can be given this kind of training.

Measurement considerations

There are good reasons to limit the number of standards against which each Year's scripts are marked. Aside from the fact that six standards at any year level is all that markers can internalise and apply at any one time, the data suggest where markers have been allowed to mark outside the given range, the numbers of students achieving those levels is small. Consideration of the standards awarded for contextual factors shows that in Year 3, only 6 per cent of students were given marks outside the range given to the markers. In Year 5, only 4 per cent of students were given marks below the range identified for Year 5 with less than 1 per cent being awarded marks above the given range. In Year 7, nearly 12 per cent of the students were marked outside of the allotted range. However, only 4 per cent of students are producing writing at the levels described by standards A, B and C with the first two standards rarely being used. This suggests that the ranges used for Years 5 and 7 need to be re-aligned but the Year 3 marking kept within the existing range.
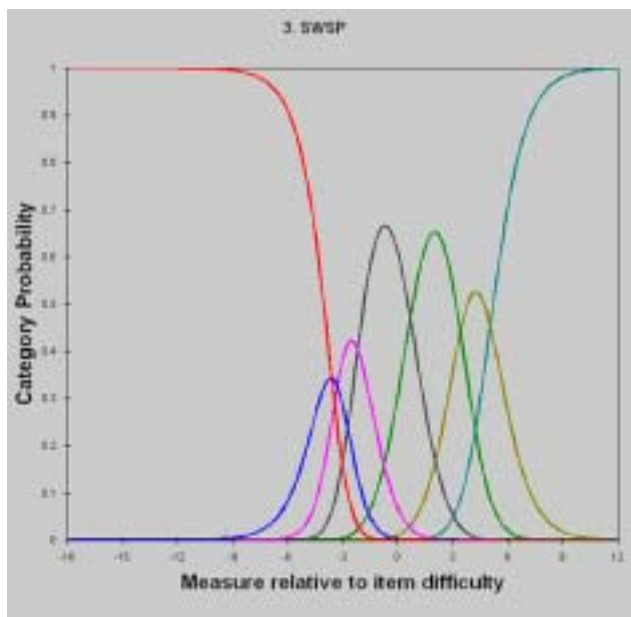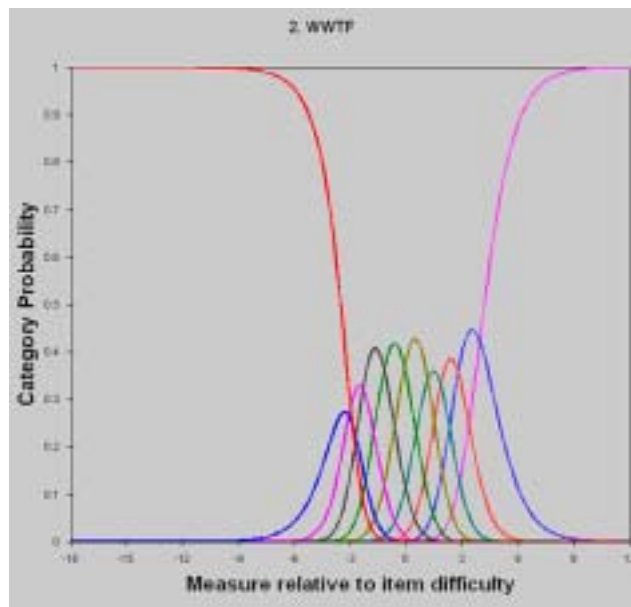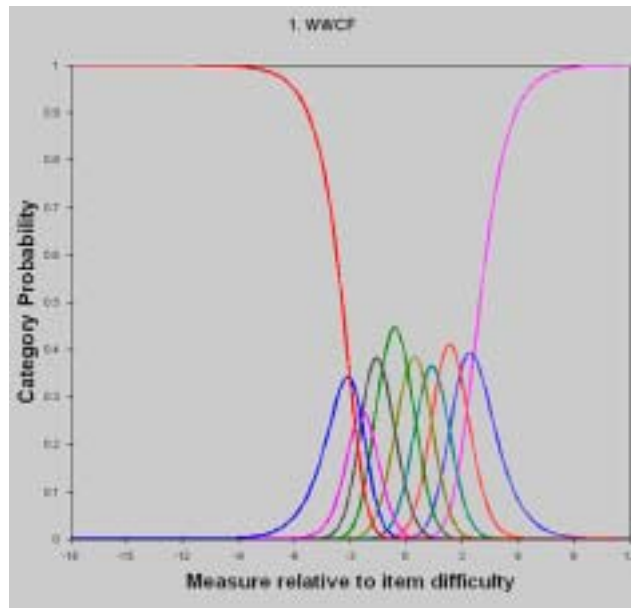
Probability curves Year 3
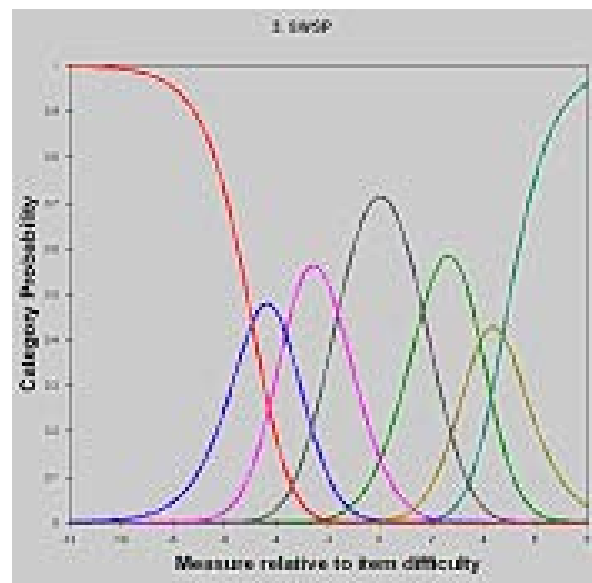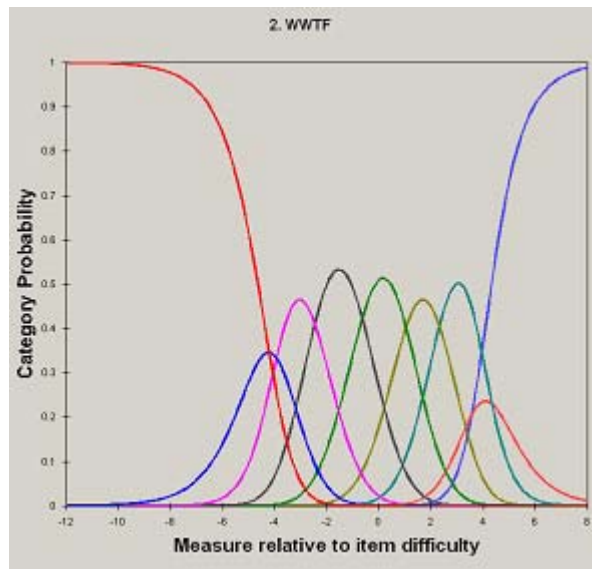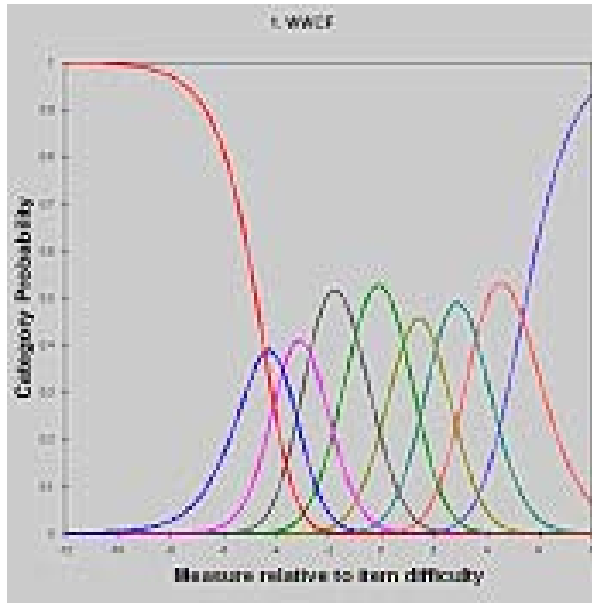
Probability curves – Year 5
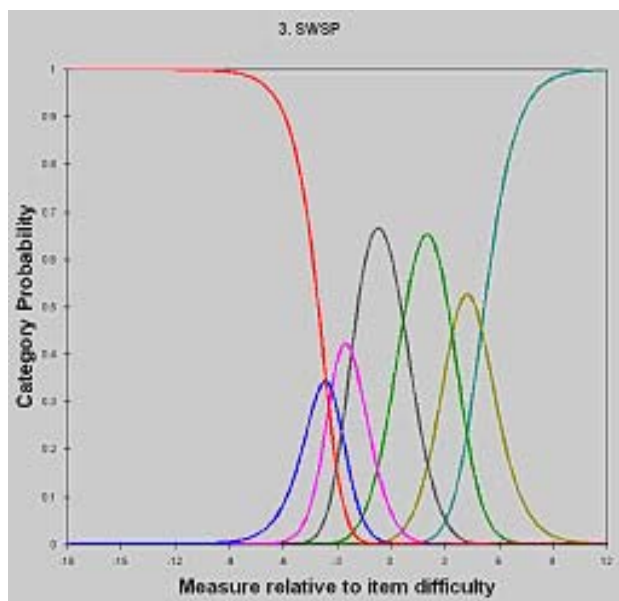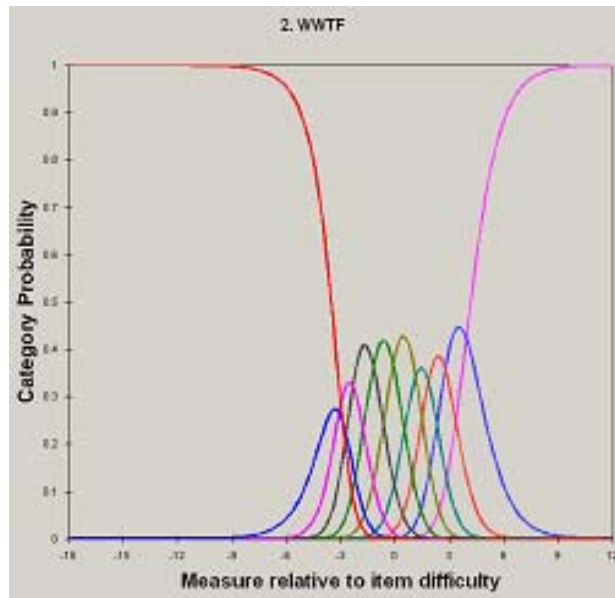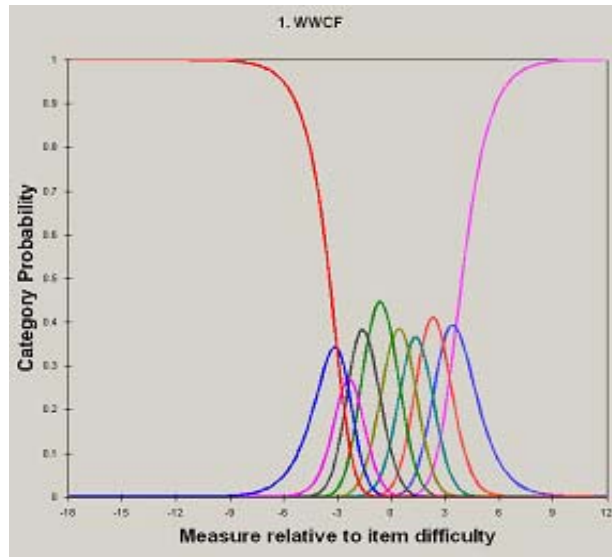
This however reduces the number of potential mark points. There is an identified need to increase the number of manageable mark points. Because many of the criteria used to assess writing are dependent variables there are two ways of increasing the number of mark points for writing. One way of doing this is to have more than one writing task. While this is probably the most educationally desirable, there are a number of significant implications for doing this. These include the cost and the impact on the students. Another way to do this is to increase the discrimination in the standards. This can be done by asking the markers to make a judgment about the standard of the writing, and then to make a second decision as to whether a student's performance is at a high, medium or low level within that standard. This will not only increase the number of mark points, it will also assist markers to make quicker judgments.

The second and perhaps additional way of increasing the number of mark points is to increase the number of criteria to three. This is particularly significant given that the writing will be reported without the spelling component. There is a need to differentiate between three levels of textual organisation - the overall level of the generic structure, the level of macro propositions, which is marked out by the lexical cohesion and the paragraphing, and the level of the clause structure. There is a difference in the way students gain control over these aspects of writing.

Using, as we did this year, a single criterion focused on structure that deals not only with the overall generic structure but also of the understanding of that and the clause structure is limiting. It is possible for students' ability to manage the clause level of writing to outstrip their ability to manage the macro-propositions. This has meant that markers had to make difficult 'on balance' decisions when scripts showed some strength in the development of clause structure while still demonstrating weaknesses in managing and ordering the macro-propositions. This was particularly the case in standards G and H. Developing two criteria to describe the textual features should allow for finer distinctions to be made about the qualities of students' writing as well as improve the efficiency of the marking operation.

In conclusion

The approach to the measurement and marking of writing adopted this year offered enough positive results to encourage us to continue with the approach and to refine it further. Writing is a complex business. Composition does not get any easier as writers develop higher and higher levels of ability. Instead, writers engage with and solve more complex problems. As test developers, the challenge is for us to engage with the complexity of the problems and identify those major milestones that are measurable and markable.

In introducing this paper, I observed that the changes we have made to the writing task have answered some questions and raised others. One of the strengths of the testing program is its capacity to raise questions—good questions about curriculum, pedagogy and assessment. The challenge, for all of us, is to answer these questions in ways that help us all develop and refine our understanding of the teaching and assessment of writing.

References

Barry, A.L. and Nielsen, D.C. (1997). *Large scale performance assessment in writing: Effects of student and teacher choice variables*. Contemporary Education; Vol 69:1, 20-27.

Bear, D.R., Invernizzi, M., Templeton, S. and Johnston, F. (2002). *Words their way; Word study for phonics, vocabulary and spelling instruction*. Prentice Hall, Upper Saddle River (NJ).

Bear, D.R. & Templeton, S. (November, 1998) *Explorations in developmental spelling: Foundations for learning and teaching phonics, spelling and vocabulary*. The Reading Teacher, Vol 52(3), 222-242.

Brighter and Scardamalia (1987). *The psychology of written composition*. Hillsdale (N.J.), Lawrence Erlbaum.

Bond, T.G. & Fox, C.M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah (N.J.), Lawrence Erlbaum.

Goldberg, G.L., Roswell, B.S. and Michaels, H. (1996). *Can assessment mirror instruction? A look at peer response and revision in a large-scale writing test*. Educational Assessment 3(4), 287-214.

Gorman, T.P., Purves, A. & Degenhart, R.E. (Eds.). (1988). *The IEA Study of Written Composition I: The International Writing Tasks and Scoring Scales*. Oxford: Pergamon Press.

Halliday, M.A.K. (1994). *An introduction to functional grammar*. (2nd Ed). Great Britain Edward Arnold.

Hunt, K.W. (1977). *Early blooming and late blooming syntactic structures*. In Charles Cooper and Lee Odell (eds), *Evaluating writing: Describing, measuring, judging. National Council of the teachers of English*, Urbana (Illinois).

Kress, G. (1994) *Learning to write*. London, Routledge, Kegan Paul.

Kress, G. & Rowan, M. (1994). *The expression of causality in children's language*. In G. Kress, *Learning to write*. London, Routledge, Kegan Paul.

Moffatt, (1981). *Active Voice*. Boynton/Cook Montclair, NJ.

Ruth, L & Murphy, S (1988). *Designing writing tasks for the assessment of writing*. Ablex, Norwood (N.J).

Templeton, S. (1992). *Theory, nature and pedagogy of higher-order orthographic development in older students*. In S. Templeton and D.R. Bear, *Development of orthographic knowledge and the Foundations of literacy: A memorial festschrift for Edmund Henderson*. Lawrence Erlbaum, Hillsdale (NJ).

Templeton, S. & Morris D. (2000) *Reconceptualizing Spelling Development and Instruction*. In Kamil, Mosenthal, Pearson & Barr, *Handbook of Reading Research III*, Lawrence Erlbaum Associates, Mahwah.

Appendix – 3, 5 and 7 marking grids

Marking grid … Schools … Year 3

Contextual factors | Textual features

	Contextual factors		Textual features	
F	a planned sequenced response relates to the topic and demands of the task attempts to engage the reader by including reasons, thoughts and actions	provides many details (some unnecessary)	opening statement to reader; statement of problem, reasons, solutions; concluding statement to the reader	basic complex sentences – time, cause, place, reason includes some thinking & feeling verbs uses modal verbs, *should have, must*
E	a planned response with an awareness of the need to build a relationship with the reader		statement of problem, reason, solutions, constructs response as a personal view – *I think you should*	stronger order/time referencing – *before that, first, later* elaborated simple sentences cause & consequence developed *as a result*
D	response to the task shows some planning and sequencing limited awareness of reader	selects ideas related to the task – may not be well integrated may have gaps	statement of problem – includes some details, statement of response	simple time referencing – *last Tuesday, then, after* simple cause relationships – *because*
C	response to task shows some planning	first person point of view	statement of problem, (action) & solution	relationships between ideas implied by sentence order and words such as "*and, then*"
B	brief response to task with little awareness of reader	meaning can be made from most of the text sounds like oral language written	statement of problem	largely simple sentences; repetitive structure everyday vocabulary some capitals and full stops used correctly
A	little organised response to task	some meaning can be made uses some unrelated ideas from stimulus	little discernible structure	partial sentences or words little or no punctuation
N	Response is unintelligible or unable to be marked.			
O	Nothing on the page.			

Spelling

	Spelling	
G	Very few spelling errors in a text that includes ambitious and complex words. Errors show an understanding of spelling/meaning principle.	*knowlidge rather than nolige*
F	Spelling of ambitious and complex words shows awareness of syllables. Correct spelling of common words with • uncommon consonant patterns • uncommon vowel patterns • predictable changes. Errors are plausible.	*behavioural, unconscious.* *ancient or column* *neighbour* *silent -silence, magic -magician* *imposabl (impossible)*
E	Shows an understanding of syllables & adding affixes. Mostly correct spelling of • multi-syllabic words • words with simple prefixes and suffixes with common conventions for addition • common homophones • common or subject-specific words with uncommon patterns. Errors may include visually similar words.	*un, re* *ment, ness, ly* *there, their, hear, here* *triangle, perimeter* *proceed for precede*
D	Shows an understanding of internal word patterns. Mostly correct spelling of • common long vowel patterns • complex consonant blends • plural or tense endings where there is no change to base words • common contractions. Errors may include complex long vowel patterns.	*train* *square, street* *wanted* *I'll* *leav, leev.*
C	Most phonemes in single syllable words are represented. Mostly correct spelling of • one syllable short-vowel words • common words with unusual spellings. Errors may include letter patterns such as *mp*, *nd* and multi-syllable words.	*dog, and* *they, have* *bup (bump)*
B	The dominant sounds within words are represented. Correct spelling of some known words.	*hafta, wonsa (once upon a time)* *a, the, I, boy, own name*
A	Words represented by letters, letter strings and some letter-like symbols.	*m = mum*
O	Nothing on the page.	

Marking grid … Schools … Year 5

Contextual factors | Textual features

H
- response is planned to take account of the relationship between the writer & the recipient
- attempts to persuade the reader to own point of view

- structures the problem and solution sequence to appeal to the reader emotionally and personally
- develops the emotional & content themes consistently
- uses simple, compound & complex sentences appropriately
- uses emotive vocabulary

G
- response shows an awareness of the formality of the relationship between reader and writer
- develops a strong personal voice
- develops subject matter from a school or class perspective

- structures the problem and solution sequence logically and factually – lacks emotive and personal appeal
- sentences contain some extending phrases
- some clustering of ideas into paragraphs
- chooses vocabulary sensitive to the relationship & purpose

F
- a planned sequenced response relates to the topic and demands of the task
- provides many details (some unnecessary)
- attempts to engage the reader by including reasons, thoughts and actions

- opening statement to reader; statement of problem, reasons, solutions; concluding statement to the reader
- basic complex sentences – time, cause, place, reason
- includes some thinking & feeling verbs
- uses modal verbs, *should have, must*

E
- a planned response with an awareness of the need to build a relationship with the reader

- statement of problem, reason, solutions, constructs response as a personal view – *I think you should*
- stronger order/time referencing – *before that, first, later*
- elaborated simple sentences
- cause & consequence developed *as a result*

D
- response to the task shows some planning and sequencing
- selects ideas related to the task – may not be well integrated
- limited awareness of reader
- may have gaps

- statement of problem – includes some details, statement of response
- simple time referencing – *last Tuesday, then, after*
- simple cause relationships – *because*

C
- response to task shows some planning
- first person point of view

- statement of problem, (action) & solution
- relationships between ideas implied by sentence order and words such as "*and, then*"

N Response is unintelligible or unable to be marked.

O Nothing on the page.

Spelling

G
- Very few spelling errors in a text that includes ambitious and complex words.
- Errors show an understanding of spelling/ meaning principle.
  - *knowlidge rather than nolige*

F
- Spelling of ambitious and complex words shows awareness of syllables.
  - *behavioural, unconscious.*
- Correct spelling of common words with
  - uncommon consonant patterns — *ancient or column*
  - uncommon vowel patterns — *neighbour*
  - predictable changes. — *silent -silence, magic -magician*
- Errors are plausible. — *imposabl (impossible)*

E
- Shows an understanding of syllables & adding affixes.
- Mostly correct spelling of
  - multi-syllabic words — *un, re*
  - words with simple prefixes and suffixes with common conventions for addition — *ment, ness, ly*
  - common homophones — *there, their, hear, here*
  - common or subject-specific words with uncommon patterns. — *triangle, perimeter*
- Errors may include visually similar words. — *proceed for precede*

D
- Shows an understanding of internal word patterns.
- Mostly correct spelling of
  - common long vowel patterns — *train*
  - complex consonant blends — *square, street*
  - plural or tense endings where there is no change to base words — *wanted*
  - common contractions. — *I'll*
- Errors may include complex long vowel patterns. — *leav, leev.*

C
- Most phonemes in single syllable words are represented.
- Mostly correct spelling of
  - one syllable short-vowel words — *dog, and*
  - common words with unusual spellings. — *they, have*
- Errors may include letter patterns such as *mp, nd* and multi-syllable words. — *bup (bump)*

B
- The dominant sounds within words are represented. — *hafta, wonsa (once upon a time)*
- Correct spelling of some known words. — *a, the, I, boy, own name*

A
- Words represented by letters, letter strings and some letter-like symbols. — *m = mum*

O Nothing on the page.

Marking grid … Schools … Year 7

Contextual factors | Textual features

	Contextual factors		Textual features
J	develops a complete & persuasive response to task tries to assert control of the writer–reader relationship	develops an authoritative voice controlled development of subject matter in a coherent argument understands & uses the recipient's point of view to persuade	deliberate manipulation of the structure for effect	links between ideas are tight & well developed begins to manipulate sentence form & length for effect increases lexical density through noun and verb groups
I	attempts to develop a relationship with the reader	shows an awareness of the values & beliefs of own community understands other points of view	elaborates the form to communicate multiple points of view	links between ideas are tight and effective elaborate ideas with *who, which, that, ing* clauses chooses vocabulary with precision punctuates to enhance meaning
H	response is planned to take account of the relationship between the writer & the recipient	attempts to persuade the reader to own point of view	structures the problem and solution sequence to appeal to the reader emotionally and personally	develops the emotional & content themes consistently uses simple, compound & complex sentences appropriately uses emotive vocabulary
G	response shows an awareness of the formality of the relationship between reader and writer	develops a strong personal voice develops subject matter from a school or class perspective	structures the problem and solution sequence logically and factually – lacks emotive and personal appeal	sentences contain some extending phrases some clustering of ideas into paragraphs chooses vocabulary sensitive to the relationship & purpose
F	a planned sequenced response relates to the topic and demands of the task attempts to engage the reader by including reasons, thoughts and actions	provides many details (some unnecessary)	opening statement to reader; statement of problem, reasons, solutions; concluding statement to the reader	basic complex sentences – time, cause, place, reason includes some thinking & feeling verbs uses modal verbs, *should have, must*
E	a planned response with an awareness of the need to build a relationship with the reader		statement of problem, reason, solutions, constructs response as a personal view – *I think you should*	stronger order/time referencing – *before that, first, later* elaborated simple sentences cause & consequence developed *as a result*
N	Response is unintelligible or unable to be marked.			
O	Nothing on the page.			

Spelling

	Spelling	
G	Very few spelling errors in a text that includes ambitious and complex words. Errors show an understanding of spelling/ meaning principle.	*knowlidge rather than nolige*
F	Spelling of ambitious and complex words shows awareness of syllables. Correct spelling of common words with • uncommon consonant patterns • uncommon vowel patterns • predictable changes. Errors are plausible.	*behavioural, unconscious.* *ancient or column* *neighbour* *silent -silence, magic -magician imposabl (impossible)*
E	Shows an understanding of syllables & adding affixes. Mostly correct spelling of • multi-syllabic words • words with simple prefixes and suffixes with common conventions for addition • common homophones • common or subject-specific words with uncommon patterns. Errors may include visually similar words.	*un, re* *ment, ness, ly* *there, their, hear, here* *triangle, perimeter* *proceed for precede*
D	Shows an understanding of internal word patterns. Mostly correct spelling of • common long vowel patterns • complex consonant blends • plural or tense endings where there is no change to base words • common contractions. Errors may include complex long vowel patterns.	*train* *square, street* *wanted* *I'll* *leav, leev.*
C	Most phonemes in single syllable words are represented. Mostly correct spelling of • one syllable short-vowel words • common words with unusual spellings. Errors may include letter patterns such as *mp*, *nd* and multi-syllable words.	*dog, and* *they, have* *bup (bump)*
B	The dominant sounds within words are represented. Correct spelling of some known words.	*hafta, wonsa (once upon a time)* *a, the, I, boy, own name*
A	Words represented by letters, letter strings and some letter-like symbols.	*m = mum*
O	Nothing on the page.	