

CHANGES IN KEY STAGE TWO WRITING FROM 1995 TO 2002

Sylvia Green, Martin Johnson, Nick O'Donovan and Pauline Sutton

A paper presented at the United Kingdom Reading Association Conference
University of Cambridge, 11 – 13 July 2003

Disclaimer

The opinions expressed in this paper are those of the authors and are not to be taken as the opinions of the University of Cambridge Local Examinations Syndicate.

Note

This research is based on data collected by the University of Cambridge Local Examinations Syndicate.

Contact Details

Sylvia Green
Head of Primary Assessment Unit
Research and Evaluation Division
University of Cambridge Local Examinations Syndicate
Syndicate Buildings
1 Hills Road
Cambridge
CB1 2EU
Tel. 01223 553844
Fax. 01223 552700
email green.s@ucles.org.uk

Introduction

Between 1995 and 2002 there have been developments in teaching and assessment intended to improve children's literacy skills. Writing has been a focus of attention because of the gap between reading and writing performance. The underachievement of boys has also been a cause for concern. In 1995 the first key stage two national tests took place and since then there has been a year on year increase in reading performance until 2002 when the results levelled out. The results for writing have shown less dramatic improvement. Literacy was brought sharply into focus with the introduction of the National Literacy Strategy (NLS) in 1998. The aim of this initiative was to raise standards in primary literacy and to do so by prescribing the content of a term by term literacy curriculum, in the framework for teaching, as well as by introducing pedagogical guidance for the structuring of the literacy hour. The fact that the key stage two English test results did not improve in 2002 has focused attention on the literacy profile once again. In the Ofsted report on the first four years of the NLS it was reported that the strategy has had a 'significant impact on standards. Nevertheless it has not been sufficiently effective to enable the government to meet its target for literacy'. A critical review of the NLS was recommended. Currently the national writing tests and mark schemes at key stages one, two and three have been changed in an attempt to increase cross key stage consistency and to provide more detailed information about writing performance.

This UCLES study comes at a time when key stage two writing is at the centre of discussion and review. The aim is to investigate changes in the narrative and discursive writing of eleven year olds from 1995 to 2002. In 1995 two writing tasks were completed by year six children, and this exercise was repeated by pupils from the same schools in 2002, using the same tasks and mark schemes.

The study explores changes in performance over time by comparing:

national curriculum levels for the two years;

a range of linguistic features;

markers' judgements;

genre effects.

One of the aims of the NLS was to broaden the range of genres in key stage two writing. In this study each child wrote a narrative and a discursive text and this enabled us to explore genre differences and to consider any changes over the intervening years.

Since 1995 teachers and markers have become better informed about the teaching of writing at word, sentence and text level, as a result of greater experience of national assessments and also the implementation of the NLS. Professional development and increased expertise may have led to changes in expectations. Any such shifts in judgements, if they have led to greater demands, could support the concept of 'expectation creep', with markers expecting higher performances for a given level. As reported by Massey et al. (2002) 'creating medium-term stability in the curricular and assessment regimes is an essential prerequisite to the maintenance of test standards.' This would have implications for comparisons of standards over time since consistency in standard setting depends on consistency in markers' judgements against a given set of criteria.

Methodology

This research was based on the writing of year six children in 1995 and 2002. In 1995 396 children each completed two writing tasks, one discursive and one narrative (see appendix 1, p.18). Fifteen schools took part in the project with a range of school types and geographical areas. The schools were not a representative sample of the nation. The aim was, however, to compare performance in the two years, a 'second order analysis' (Goldstein and Heath, 1999; Pollitt, 2001) and for this it is important that the two samples be well matched so as to cancel out as many irrelevant factors as possible. In 2002 the exercise was replicated with 540 year six children from the same schools. A combination of marking strategies was used to gather a range of detailed information about changes over time.

In Phase One the following were compared to show changes over time in holistic levels for the whole sample:

- 1995 narrative and discursive **original holistic levels** awarded in 1995
- and
- 2002 narrative and discursive **original holistic levels** awarded in 2002

In Phase Two, based on a sub-sample of 1995 and 2002 scripts matched by level, the following variables were compared:

- narrative typed **re-mark holistic level** to investigate changes in markers' judgements over time plus the effect that typing had on judgements
- discursive hand-written **re-mark holistic level** to investigate changes in markers' judgements over time, allowing the 'typing effect' to be isolated
- typed narrative and hand-written discursive **analytical marking** for appropriacy, ideas and expression, using grades and codes to investigate whether features of writing within levels had changed over time

Phase One marking (original holistic levels)

Each piece of writing was awarded a level, using the national performance descriptions, with each level subdivided into a ten-point scale to allow greater discrimination across the range of performances. Experienced key stage two markers had carried out the marking in 1995 and some of the original 1995 markers also marked scripts in 2002. Co-ordination exercises took place in both years during which co-ordination scripts were marked and judgements were discussed. Some of the 1995 co-ordination scripts were also included during the 2002 co-ordination exercise. It was important to find out if any changes over time were real changes in writing performance or whether they were the result of changes in markers' judgements.

Phase Two marking

In phase two the issue of changes in markers' judgements over time was addressed. A sub-sample of scripts was given re-mark holistic levels to find out if they were judged differently. Scripts from 1995 were selected to represent the range of performances within the three main levels, before being matched with scripts from similar level points from 2002 (see Table 1 below). The scripts from the two years were then mixed before

being sent to markers to ensure that they would not know the original mark or date of the scripts.

TABLE 1 – Sub-sample for re-mark exercise

	Narrative ('Gate' task)		Discursive ('Tiger' task)		Total
	1995	2002	1995	2002	
Level 3	20	20	20	20	80
Level 4	20	20	20	20	80
Level 5	20	20	12	12	64
Total	60	60	52	52	224

The narrative scripts from 1995 were not available; however, they had been saved on computer verbatim. It would have been unfair to directly compare levels awarded to typed and hand-written scripts since any changes could have been the result of 'typing effects'. Consequently it was decided to type up the 2002 narrative scripts to ensure that valid comparisons could be made. The 1995 and 2002 hand-written discursive scripts were available and it was decided that they should be re-levelled in their original form.

As well as the re-marking with holistic levels, the writing was marked analytically to investigate detailed changes in the writing within the levels. The analysis of the writing was based on the three components of communicative competence identified by Canale and Swain (1980) and used in The English Language Skills (TELS) Profile (1987):

appropriacy of register, structure and style to the defined task and audience
(sociolinguistic competence);

ideas the selection and organisation of information in a way which best meets the requirements of the task (discourse competence);

expression in terms of structuring, paragraphing, syntax, punctuation and spelling (grammatical competence).

Appropriacy and ideas were graded according to sets of descriptors from grade 1 to grade 4 (see appendices 2 and 3, pp.19, 20). Expression was analysed using a combination of quantitative coding (see appendix 4, pp.21, 22) and descriptors (see

appendix 5, p.23). The coding frameworks were based on those used for the analyses of pupil performance designed by the Qualifications and Curriculum Authority (QCA). The following data were also collected for each child: school; gender; date of birth; teacher assessment level for writing; national test level for writing (available in 2002 only).

Markers' judgements

As well as comparing original holistic levels and re-mark holistic levels to see if markers' judgements had changed over time we were able to compare 1995 and 2002 markers' reports on the factors that had affected their judgements and thereby explore how their perceptions had changed over time. By investigating changes in judgements it was possible to validate apparent changes in children's writing performance.

Impact of typed scripts on markers' judgements

Circumstances dictated that the sub-sample of 1995 narrative scripts selected for phase two was typed. Consequently, the matched 2002 hand-written scripts were also typed. It was necessary to estimate a 'typing effect' (i.e. the effect that typing had on markers' judgements). This was done using the 2002 sub-sample. The difference between the original holistic levels on the hand-written scripts and the re-mark holistic levels on the typed versions was calculated. Once we knew the extent of the typing effect we were able to take it into account when considering whether markers' judgements had changed over time. Given the current debate about computer mediated assessment, this was an interesting additional issue.

Findings

Phase One marking (original holistic levels)

For each task the results from 1995 and 2002 were compared. (See Table 2)

TABLE 2 – Results of Phase One marking

Table 2a - Discursive task

YEAR	GENDER	Mean	Std. Deviation
1995	Girl	3.54	0.79
	Boy	3.24	0.84
	Total	3.37	0.83
2002	Girl	4.08	0.76
	Boy	3.85	0.78
	Total	3.96	0.78

Table 2b - Narrative task

YEAR	GENDER	Mean	Std. Deviation
1995	Girl	4.03	1.00
	Boy	3.73	0.90
	Total	3.87	0.96
2002	Girl	4.25	0.91
	Boy	3.88	0.92
	Total	4.06	0.94

There was an overall improvement in performance from 1995 to 2002 with greater improvement in discursive than in narrative writing. Analysis of variance tests showed that the improvement in discursive levels over time was substantial and significant, while the changes in narrative levels was statistically significant but less substantial (see Table 3, appendix 6, p.24). Although most schools showed improvement over time, the scores for one school had decreased in both tasks by half a level, and in a minority of schools there were some decreases in scores (see Table 4, appendix 6, p.25). This is the source the interaction effect noted in the results of the analysis of variance.

The mean scores for girls were consistently higher than for boys; standard deviations for the different genders and years were similar, though variation in scores for the narrative task was, on the whole, higher.

The relationships between the levels awarded for each task and teacher assessment levels were investigated for both years. National test levels were only available for 2002 so this was included in the 2002 comparisons (see Tables 5 and 6, appendix 6, p.25).

Teacher assessments correlated better with the narrative than with the discursive task and there was no real change in the relationships over time. The 2002 data indicated that the discursive task and the national test results correlated to a lesser degree with the teacher assessment levels than the narrative task did. Overall, even with a time lag between the assessments and the differential motivational effects, the correlation between these measures is relatively close. The evidence therefore suggests that the four performance indicators do bolster each other and cover similar domains, thereby lending weight to the validity of the observations made.

Phase Two marking

The aim of the analytical marking was to investigate qualitatively how writing had changed over time by investigating specific features (see coding frame, appendix 4, pp.21, 22). For this purpose a sub-sample was drawn for each year, with 20 scripts for each level (apart from the discursive scripts for level 5, as only 12 children had achieved that level in 1995. See Table 1, p.5). It is important to bear in mind that the sub-samples of scripts from each year were matched according to national curriculum levels and that any changes detected within the sub-sample indicate the way writing awarded a particular level in 1995 differs from writing awarded the same level in 2002.

The general trend was that the writing *within* the levels had improved between 1995 and 2002. There were some exceptions and these will be discussed later. It should be noted that these improvements are in addition to changes over time in the number of pupils achieving each level as reported from the marking in phase one.

Discursive task

When the scripts were graded for appropriacy and ideas, those with original holistic level 3 in 2002 did less well than those originally judged as level 3 in 1995. Scripts with original holistic level 5 in 2002 were more successful in ideas and appropriacy than those originally judged as level 5 in 1995. The trend seemed to be that the more able had improved over time while the less able were judged to have got worse in these aspects of writing (see Tables 7a & 7b, appendix 6, p.26).

Sentence co-ordination was deemed more effective in 1995 scripts while subordination was deemed more effective in 2002 scripts suggesting an improvement in the complexity of the writing at sentence level.

The coding data for sentence demarcation and 'internal punctuation' were merged into one category due to the infrequency of such errors. There were fewer errors overall in 2002 texts, although this was marginal, and at level 5 there were slightly more errors in the 2002 writing. There were also more comma splice errors and more frequent use of clausal and list commas in 2002 texts. The coding for punctuation included quantitative data from a 100 word extract as well as an overall judgement for the whole text. Interestingly, the 2002 level 5 scripts had higher overall grades for punctuation. This

could suggest that counting errors within the 100 word extracts was telling only part of the story and that over the whole text the errors could belie greater sophistication.

Textual coherence was judged to be better in 1995 writing with little change in the effectiveness of openings and an improvement in endings in 2002 scripts.

Within the context of overall improvement and improvements in a range of features of writing, the evidence suggested that there was an increase in the number of spelling errors. This increase was not statistically significant and the number of errors was small in the context of the 100 word extracts from which they were taken (see Table 8, appendix 6, p.26).

Narrative task

In most areas the pattern of the changes was similar to those for the discursive task. In appropriacy and ideas there was little change among less able writers while, as for the discursive task, there was improvement among more able writers (see Tables 9a & 9b, appendix 6, p.27).

Sentence co-ordination and subordination were more effective in 2002, again indicating improvements in the complexity of sentence structure.

The quantitative punctuation coding indicated that there were more sentence demarcation errors in 1995 with few 'internal punctuation' errors in either year and those that were found were erratically dispersed. As in the discursive task, the overall grades for punctuation in the 2002 scripts were higher than those for the 1995 scripts at level 5, even though there were more errors in the 100 word extracts. Although there were more comma splice errors in 2002 overall, this was because of the exceptionally high number amongst the level 3 scripts and there were fewer at levels 4 and 5.

Unlike in the discursive task, textual coherence and openings were better in 2002, with endings better at levels 3 and 4.

The spelling followed a similar pattern as for the discursive task and changes were not statistically significant (see Table 10, appendix 6, p.27).

The other features included in the coding showed little or no change over time.

Typing effect

The re-mark holistic levels for the 2002 narrative scripts were influenced by a typing effect. The evidence suggested that markers' judgements were more severe when they marked a typed version of a script, with a decrease of about half a level. The standard deviation for the re-mark holistic levels was similar (see Table 11, appendix 6, p.28).

Markers' judgements – changes over time

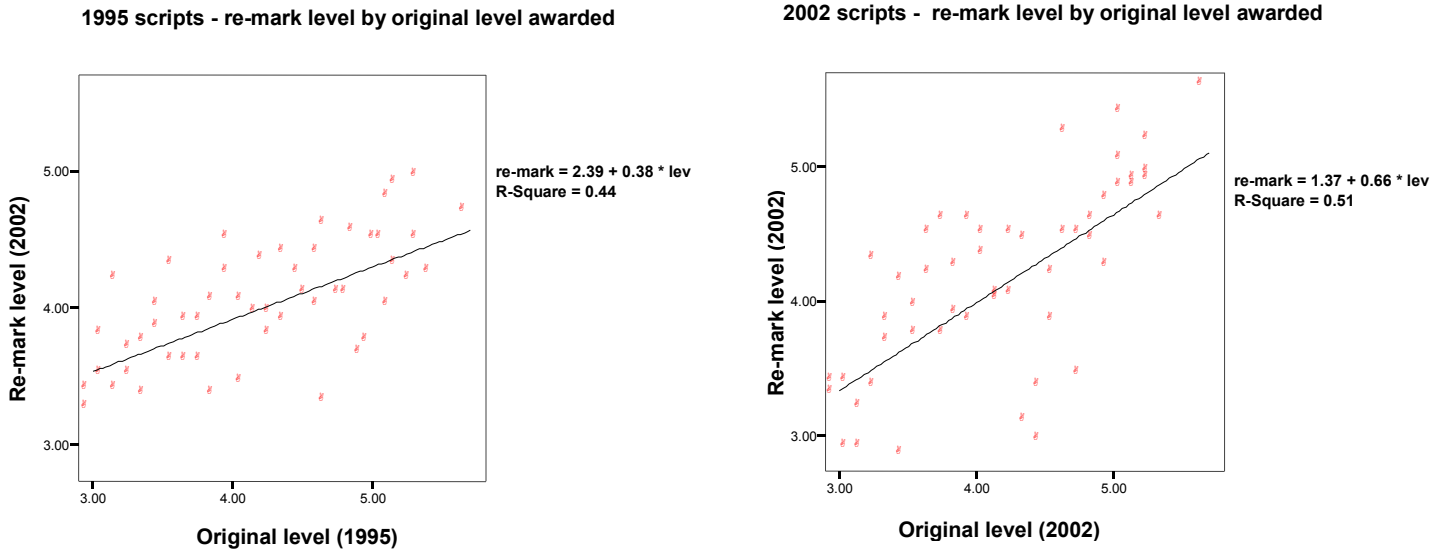
Having investigated the typing effect, we were able to consider how far differences between original and re-mark holistic levels could be related to changes in markers' judgements over time.

Discursive task

The 1995 scripts with original holistic levels 3 and 4 achieved almost identical re-mark holistic levels to their 2002 counterparts (see Table 12, appendix 6, p.28). Since the two samples had been matched by level this was further evidence that judgements had remained fairly consistent over time. However, this was not the case at level 5 where the mean re-mark holistic level for 2002 scripts was higher, by half a level, than the mean re-mark holistic level for the 1995 scripts which also had original holistic level 5. It should be noted that at level 5 only 12 scripts were compared from each year. The standard deviation for these re-levelled scripts was higher for 2002 scripts, while for the 1995 scripts it was almost identical.

When original discursive levels were plotted against re-mark levels for the 1995 and the 2002 scripts (see Figure 3 below) the changes over time noted earlier can be seen resulting from the stricter marking at level 5.

FIGURE 3 – Discursive task: Changes over time in relationship between original level and re-mark level



The findings indicate that, since 1995, there have been some changes in judgements as described, but that overall, at levels 3 and 4 especially, judgements have remained similar. Where there have been changes, the demands have increased and, if anything, it has become more difficult to gain level 5.

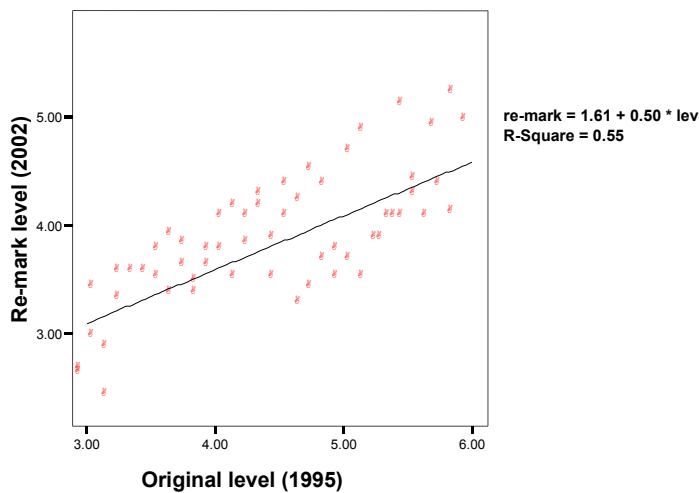
Narrative task

It appeared, from the investigation into 'typing effect', that typed responses were judged more harshly than those which were hand-written. The mean re-mark holistic levels for 1995 scripts originally judged as levels 3 and 4, were only marginally higher than those for the 2002 scripts. However, the mean re-mark holistic level for 2002 scripts originally judged as level 5, was half a level higher than for those judged to be level 5 in 1995. The level by level standard deviations of re-mark levels were relatively similar for both years. However, the overall standard deviation for the 2002 script re-mark levels was .23 of a level higher than for the 1995 script re-mark levels, suggesting better discrimination in the 2002 scripts (see Table 13, appendix 6, p.29).

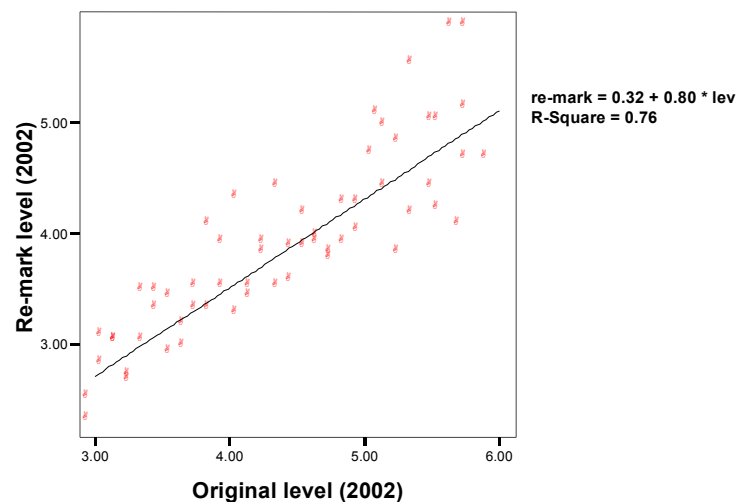
The scatter-plots for the narrative task present a similar picture to those for the discursive task, bearing in mind the typing effect which led to a lowering of re-mark levels (see Figure 4).

FIGURE 4 – Narrative task: Changes over time in relationship between original level and re-mark level

1995 scripts - re-mark level by original level awarded



2002 scripts - re-mark level by original level



The evidence suggested some change in marker judgements as overall the 2002 scripts gained slightly higher re-mark levels, relative to their original levels, than the 1995 scripts. As the overall mean indicates, the difference was marginal. However, as for the discursive task, the difference was greater for scripts with original holistic level 5, by a near identical margin of half a level. This compensated for the fact that 1995 scripts originally assessed as levels 3 and 4 performed slightly better than their 2002 counterparts in terms of re-mark holistic levels. There was a high mark re-mark correlation for the 2002 scripts (.87) while the correlation for the 1995 scripts was lower (.74). This was still a strong correlation; nevertheless the difference between correlations could indicate some change in markers' judgements between 1995 and 2002.

Markers' reports

The markers in 2002 reported on the features of the writing that they felt were more and less significant as they made their judgements at levels 3 and 5.

Level 3

In 1995 markers found it more difficult to define and identify non-technical features of writing. Markers in 2002 described and defined salient features in more detail, whereas in 1995 comments tended to be more vague and general. There was a feeling that markers in 1995 viewed the marking process in a 'kinder' way in that they used terms such as 'overlooking errors' and 'not expecting errors to be sustained'. The language suggested a more positive perspective and a more flexible approach. The standards implied by the detail given in the 2002 markers' reports suggested that 'things had got tougher', although, it is possible that markers could have applied the criteria in a similar way in both years, without verbalising the details.

Level 5

There were more technical and word/sentence level features specified by 2002 markers. Features mentioned in the 1995 reports also appeared in 2002, but there were more added in 2002. 'Maturity' was mentioned in 1995, but it was a vague concept, whereas in 2002 more detailed features were listed which could be described as 'features of more mature writing'. Markers were also asked to suggest criteria of their own and when they did this for level 5 they expanded on their original comments resulting in increased overlap between the years. This could suggest that they were more comfortable expressing their expectations in their own terms, rather than commenting on the criteria given for marking. This was especially the case for the narrative writing. Overall, as at level 3, the markers' comments suggested that it had become more difficult to achieve level 5.

Discussion

The evidence from the study showed that there had been improvements in children's writing between 1995 and 2002. However, this general statement masks a range of different issues within the overall picture of improvement. These differences relate to genre, level, expectations and various aspects of writing.

It was clear from the mean original holistic levels for the whole sample that there had been significant improvements in both narrative and discursive writing. However, changes were more significant and substantial in the latter, possibly because of the NLS emphasis on a range of genres during the primary phase. When we add to this the evidence that markers' judgements have been fairly consistent over time, especially at levels 3 and 4, then the improvements noted can be taken seriously. In fact, closer analysis suggested that at level 5 there had been 'expectation creep', with markers in 2002 demanding more for that level. This, together with evidence that standards have stayed fairly consistent at levels 3 and 4, supports the case that writing performance has improved over time.

In the substantive areas of ideas and appropriacy there were clear indications that scripts awarded level 5 in 2002 gained higher re-mark levels than those matched scripts from 1995. It should be noted that this could result from 'different' criteria being applied for these aspects of writing in 2002 than would have been applied in 1995. However, the comparison of comments made in markers' reports suggested that, if anything, 'things had got tougher' in 2002. The fact that the quantitative data indicates that the level 5 standard has shifted, adds to the evidence from the markers that performance has improved over time and in the light of the evidence, it could be argued that improvements have been underestimated at level 5. The improvements will be encouraging for teachers, children and those responsible for new initiatives implemented to achieve this aim. However, any shift in markers' judgements against assessment criteria has implications for comparisons of standards over time. It could be argued that it is inevitable that teachers' and markers' judgements will change to some extent and it is surprising that they have remained relatively consistent at levels 3 and 4. There was also evidence that markers in 2002 were more able to differentiate between the 2002 scripts than between those from 1995 possibly because of more clearly defined expectations of teachers and markers.

The scripts selected for detailed coding were matched by national curriculum level and so any differences indicate how the writing has changed at each of the levels, as awarded in 1995 and 2002. The changes in levels awarded to the full sample indicated that there had been significant improvements in performance in both tasks, (particularly in discursive) and so the differences indicated by the coding need to be set in the

context of overall improvements in national curriculum levels achieved. Although the coding suggested that, overall, the 2002 scripts were better than their 1995 counterparts, in some areas there was little or no change and in some 1995 scripts were better.

Overall, sentence structure and punctuation had improved over time although, in the discursive task, there were more punctuation errors at level 5 in 2002. There were also more clausal and list commas used. There were more comma splice errors in 2002, although in the narrative task this was because of a high number concentrated in level 3 scripts. Punctuation generally had improved and the increased frequency of comma splices may be a result of children trying to write more complex sentences but not being able to combine this with the appropriate complexity of punctuation. This appeared to be the case for the less able in narrative writing. There was some evidence to suggest that spelling was better in 1995 but the differences were not statistically significant and the numbers of errors involved were small.

The picture for coherence was different for the two tasks, with improvements in the narrative but not in the discursive writing. Openings and endings at level 5 were also better in narrative in 2002 while discursive openings had changed little and discursive endings were better overall. From these mixed results it seems that these aspects of writing have improved more in narrative than in discursive writing. This could be because textual features of narrative are more familiar and the introduction of a wider range of genres with their associated forms may be more difficult to consolidate.

Although the effect of typing on markers' judgements was an accidental theme, it proved to be an interesting side issue. It was clear that markers' judgements did change when a script was typed, with typed scripts marked down by half a level compared to their hand-written counterparts. It could be that typing removes the temptation for the marker to make allowances for less able or less mature writers. It may be that markers react more sympathetically when they see scripts hand-written by children who are clearly struggling to present their ideas, possibly making allowances and applying the criteria differently. Also, the clearer presentation of the typed script may accentuate the errors in the writing and markers may have their own 'higher expectations' of a piece of typed text. The evidence of 'different standards' has implications for any future discussions

about assessment of typed scripts since such inconsistency would need to be researched further and taken into account when comparing performance over time.

Overall, the findings provide some good news and are encouraging for those who aim to improve performance and to maintain standards over time. Performance has improved in terms of the number of children achieving given levels in this study, as in the national statistics. As well as this change, there is also evidence to suggest that some of the criteria used in the judgements are applied more severely than they were. Although the evidence from the study suggests that there have been significant improvements in children's writing over time, the more detailed analysis poses questions about how writing at particular levels has changed. Overall, children achieved higher levels in 2002 than in 1995 and not only did more children achieve level 5, but more sophisticated writing was expected at this level.

The findings from this research raise issues related to features of writing, the impact of new initiatives and the way that criteria are applied. Changes in these areas have implications for teaching and assessment and are important for appropriately targeted teaching and for reliable assessment.

References

Canale, M. and Swain, M. (1980) 'Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing' *Applied Linguistics*, 1.1 pp 1 - 47

Department for Education and Employment (1998) *The National Literacy Strategy: Framework for Teaching*. London: DfEE

Goldstein, H. and Heath, A. (eds) (2000) *Educational Standards*. p 8. Oxford: Oxford University Press, for The British Academy

Hutchinson, C. and Pollitt, A, with Lillian Munro (1987) *The English Language Skills Profile*. Godfrey Thomson Unit, University of Edinburgh: Macmillan Education

Massey, Alf, Green, Sylvia, Dexter, Trevor and Hamnett, Lisa (2002) *Comparability of national tests over time: KS1, KS2 and KS3 standards between 1996 and 2001. Final report of the Comparability Over Time Project*. London: Qualifications and Curriculum Authority

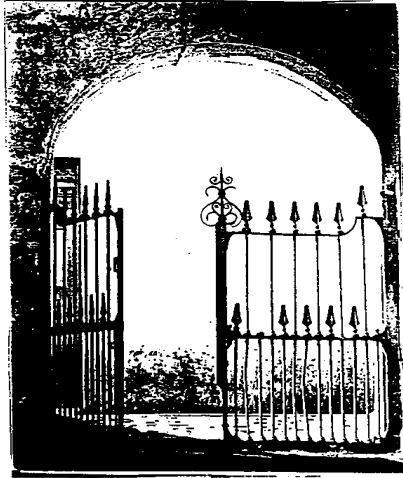
Office for Standards in Education (2002) *National Literacy Strategy: the first four years*. HMI No.55. London: Ofsted

Pollitt, A. (2001) 'Educational Standards': [review essay] *Assessment in Education*, 8.2 pp 261 - 271

Narrative Task

"The gate was always locked. But on that day someone had left it open"

- Write a story about what happened when you went through the gate.



Discursive Task

The Tiger

The tiger behind the bars of his cage growls,
The tiger behind the bars of his cage snarls,
The tiger behind the bars of his cage roars.

Then he thinks,
It would be nice not to be behind bars all
The time

Because they spoil my view
I wish I were wild, not on show.
But if I were wild, hunters might shoot me,
But if I were wild, food might poison me,
But if I were wild, water might drown me.

Then he stops thinking
And
The tiger behind the bars of his cage growls,
The tiger behind the bars of his cage snarls,
The tiger behind the bars of his cage roars.

Peter Niblett



Read the poem carefully.

Should animals be kept in zoos?

Write about why it may be a good thing to keep animals in zoos and why it may be a bad thing. Be sure to tell us what you think.

Grade Descriptors

Appropriacy

- S1** **In narrative** the reader is not engaged due to lack of plot development and characterisation. Characters are introduced but not described and events are simply listed. Any attempts at description or dialogue are limited and ineffective. Simple spoken language is used with many sentences starting the same way. Some story language is used such as, Suddenly....., One day..... evidence of the beginning of written style.
In discursive the purpose may not be completely clear. A point of view is stated without evidence, discussion or logical argument.
- S2** **In narrative** the writer includes some extra detail about characters, perhaps including dialogue/reported speech. The writing is more lively and the writer attempts to interest the reader by commenting on characters and events, perhaps with some simple description and expanded phrases.
In discursive the writer signals a discussion at least one point on each side with a final statement
- S3** **In narrative** the introduction is effective, establishing a context and engaging the reader. There may be different narrative techniques e.g. opening with action/dialogue, time/place shifts. Appropriate choices are made e.g. use of SE and colloquialisms. Register is sustained throughout. Interest is engaged through the use of effective language, creation of tension/atmosphere, although not necessarily sustained throughout.
In discursive A series of points is made with some evidence or justification. There may be a direct appeal to the reader. An appropriate level of formality is used and sustained.
- S4** **In narrative** the reader's attention is engaged and sustained throughout. Sophisticated, mature language is confident and controlled. The features of narrative e.g. linear time lines, conflict, relationships are confidently managed. A range of sentences and vocabulary is used to create effect for the type of story. Literary features such as alliteration, figurative language or deliberate use of dialect for characterisation may be used. Descriptions of feelings add to overall interest.
In discursive points are made in a balanced way and are expanded and supported with evidence and comment. Reasoning connectives and emotive words make the case strongly for the reader. The final statement echoes previous points.

Grade Descriptors

Ideas

- S1** The writing relates to the theme but without enough organisation to suit the purpose. There is no satisfactory conclusion.
In narrative there are some basic elements of story structure – an opening, more than one character and at least two events in chronological sequence.
In discursive the writing has at least two statements but without elaboration. Some basic organisational features may be present, such as – a heading or introductory or concluding phrase.
- S2** There is some logical organisation of ideas to suit the purpose but ideas are not fully integrated. There is a conclusion but it does not round off or summarise effectively.
In narrative events are organised into beginning, middle and simple ending. Events are related to one another though not necessarily well paced.
In discursive an introductory statement is used, followed by a series of points in a sensible order. Coverage of individual ideas may be too detailed or too brief leading to overall uneven coverage.
- S3** Ideas are selected, organised, developed and sustained. There is a conclusion which rounds off or summarises the ideas well.
In narrative the writing is coherent and well paced. Beginning, middle and ending can be distinguished and all events are logically related.
In discursive The writing is coherent and balanced, covering a range of aspects of the subject matter. Ideas are sustained and developed in a logical way. The writing is structured with relevant information, a series of relevant points and a suitable concluding phrase or sentence.
- S4** Ideas are selected and organised effectively. The piece is complete and there is evidence that there has been forethought and planning.
In narrative the writing is well constructed. Dialogue, action and description are interwoven and there is development of the theme / plot. Paragraphing may be used to mark main divisions of the narrative such as – opening, main events and ending. The sequence of events is confidently managed.
In discursive the writing is well structured and well balanced with good coverage of a range of well ordered points. The introduction establishes the context and the series of expanded points links to the purpose of the writing. There is a considered ending giving the piece a sense of effective completion. The range of information is covered in appropriate detail and the organisation may be supported by paragraphing.

Appendix 4
CHANGES OVER TIME CODING FRAME

Script number	
---------------	--

Appropriacy

S

Ideas

S

Expression Coding Frame

Word count	Word length	Flesch reading ease	Flesch reading grade level

Sentence structure

T- units

S

100 word extract	Number			
Number of sentences (defined by grammatical structure)				
Number of finite verbs				
Co-ordinated clauses				
Effectiveness rating	1	2	3	4
Subordinate clauses				
Effectiveness rating	1	2	3	4

Sentence demarcation

Whole text judgement

S

100 word extract	Number correct	Number incorrect	Number omitted
Capital letter to begin sentence			
Full stop to end sentence			
Number of comma splices			

Appendix 4 cont.

Internal punctuation

Whole text judgement

S

100 word extract	Number correct	Number incorrect	Number omitted
Commas to separate items in a list			
Commas to demarcate clauses			

Paragraphs

Whole text	Number
Total number of paragraph breaks	
Number of links between paragraphs	

Coherence

Opening

S

 ending

S

 coherence

S

Grammar

Whole text	Number of errors
Verb agreement	
Tenses	
Pronouns	

Spelling

100 word extract	Number
Numbers of errors	
Range and sophistication of vocabulary (3: most sophisticated.)	1 2 3

Word class

100 word extract	Number of different words in class	Number of words used (including repeats)
Adjectives		
Adverbs		

Sentence Demarcation

- S1 The writer shows some awareness of sentence demarcation.
- S2 Sentences correctly demarcated by capital letters and full stops less than half the time
- S3 Sentence demarcation correct most of the time
- S4 Sentence demarcation correct and accurate, one error acceptable

Internal Punctuation

- S1 Internal sentence punctuation omitted
- S2 Beginning to use punctuation within sentences – sometimes correct
- S3 A range of punctuation including , ‘ “ “ within sentences – usually correct
- S4 Correct punctuation used accurately to clarify meaning, one error acceptable

Opening

- S1 Genre/theme are signalled, limited ideas on character, setting, time-frame (at least one included)
- S2 Detailed interesting ideas about character, setting, and time-frame are given
- S3 The opening engages the reader's attention with creative strategies, language, ideas.

Ending

- S1 No closure or closure which is incongruent
- S2 Adequate closure with completion
- S3 An ending which draws the story to an effective ending, ties loose ends and echoes the earlier content

Coherence

- S1 Difficult to follow
- S2 Logical thread which reaches a conclusion
- S3 Sustained coherence in a more complex text

Appendix 6
Statistics

TABLE 3 – Analysis of test score variance

Table 3a – Tests of between-subjects effects: Discursive

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	218.55	59	3.70	7.58	0.00
Intercept	9952.30	1	9952.30	20355.98	0.00
YEAR	65.22	1	65.22	133.39	0.00
GENDER	6.80	1	6.80	13.91	0.00
SCH_NO	60.31	14	4.31	8.81	0.00
YEAR * GENDER	0.07	1	0.07	0.15	0.70
YEAR * SCH_NO	56.11	14	4.01	8.20	0.00
GENDER * SCH_NO	4.41	14	0.32	0.64	0.83
YEAR * GENDER * SCH_NO	5.21	14	0.37	0.76	0.71
Error	405.31	829	0.49		
Total	12944.58	889			
Corrected Total	623.86	888			

Table 3b – Tests of between-subjects effects: Narrative

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	198.48	59	3.36	4.57	0.00
Intercept	12474.35	1	12474.35	16942.47	0.00
YEAR	9.19	1	9.19	12.48	0.00
GENDER	15.37	1	15.37	20.87	0.00
SCH_NO	67.17	14	4.80	6.52	0.00
YEAR * GENDER	0.59	1	0.59	0.81	0.37
YEAR * SCH_NO	71.16	14	5.08	6.90	0.00
GENDER * SCH_NO	10.08	14	0.72	0.98	0.48
YEAR * GENDER * SCH_NO	15.64	14	1.12	1.52	0.10
Error	639.83	869	0.74		
Total	15579.76	929			
Corrected Total	838.30	928			

TABLE 4 – Changes over time by school

School number	Change in mean level from 1995 to 2002 (narrative)	Change in mean level from 1995 to 2002 (discursive)
1	0.22	0.98
2	-0.17	0.56
3	0.05	-0.24
4	0.51	1.02
5	-0.51	-0.52
6	0.09	0.83
7	0.57	0.69
8	0.73	1.35
9	-0.96	-0.04
10	1.31	0.97
11	0.81	1.11
12	0.91	1.12
13	-0.49	-0.01
14	-0.06	0.79
15	-0.32	0.24

TABLE 5 – Correlations between teacher assessment and task performance 1995

		Teacher assessment		
Teacher assessment	Pearson Correlation	1.00		
	Sig. (2-tailed)	.		
	N	397	Discursive level	
Discursive level	Pearson Correlation	0.59	1.00	
	Sig. (2-tailed)	0.00	.	
	N	366	370	Narrative level
Narrative level	Pearson Correlation	0.69	0.67	1.00
	Sig. (2-tailed)	0.00	0.00	.
	N	388	360	390

All correlations are significant at the 0.01 level (2-tailed).

TABLE 6 – Correlations between teacher assessment, task performance and national test levels 2002

		Teacher assessment			
Teacher assessment	Pearson Correlation	1.00			
	Sig. (2-tailed)	.			
	N	544	Discursive level		
Discursive level	Pearson Correlation	0.63	1.00		
	Sig. (2-tailed)	0.00	.		
	N	515	519	Narrative level	
Narrative level	Pearson Correlation	0.71	0.61	1.00	
	Sig. (2-tailed)	0.00	0.00	.	
	N	539	510	539	National test level
National test level	Pearson Correlation	0.63	0.57	0.66	1.00
	Sig. (2-tailed)	0.00	0.00	0.00	.
	N	532	503	532	532

All correlations are significant at the 0.01 level (2-tailed).

TABLE 7 – Discursive task crosstabulations

Table 7a - Appropriacy

Original level attributed * Discursive appropriacy * YEAR Crosstabulation

Count			Discursive appropriacy				Total
YEAR			1	2	3	4	
1995	Original level	3	7	9	4		20
	attributed	4	1	10	8	1	20
		5		3	5	4	12
	Total		8	22	17	5	52
2002	Original level	3	11	3	5	1	20
	attributed	4	4	10	4	2	20
		5	1		5	6	12
	Total		16	13	14	9	52

Table 7b - Ideas

Original level attributed * Discursive ideas * YEAR Crosstabulation

Count			Discursive ideas				Total
YEAR			1	2	3	4	
1995	Original level	3	9	9	2		20
	attributed	4	1	13	6		20
		5		4	4	4	12
	Total		10	26	12	4	52
2002	Original level	3	11	7	1	1	20
	attributed	4	5	12	1	2	20
		5		1	6	5	12
	Total		16	20	8	8	52

TABLE 8 – Discursive task spelling errors (per 100 word extract)

Mean		
Original level attributed	Discursive spelling errors	
	YEAR	
	1995	2002
3	5.55	6.45
4	2.65	3.80
5	2.50	3.08
Total	3.73	4.65

(ANOVA tests indicate that Year is not significantly related to number of spelling errors)

TABLE 9 – Narrative task communicative competence crosstabulations

Table 9a - Appropriacy

Original level attributed * Narrative appropriacy * YEAR Crosstabulation

Count			Narrative appropriacy				Total
YEAR			1	2	3	4	
1995	Original level	3	15	3	2		20
	attributed	4	1	14	5		20
		5	1	6	9	3	19
	Total		17	23	16	3	59
2002	Original level	3	14	5	1		20
	attributed	4	3	8	8	1	20
		5	1	2	8	9	20
	Total		18	15	17	10	60

Table 9b – Ideas

Original level attributed * Narrative ideas * YEAR Crosstabulation

Count			Narrative ideas				Total
YEAR			1	2	3	4	
1995	Original level	3	12	8			20
	attributed	4	3	12	5		20
		5	2	12	2	3	19
	Total		17	32	7	3	59
2002	Original level	3	12	6	2		20
	attributed	4	5	11	3	1	20
		5	1	5	6	8	20
	Total		18	22	11	9	60

TABLE 10 – Narrative task spelling errors (per 100 word extract)

Mean		
Original level attributed	Narrative spelling errors	
	YEAR	
	1995	2002
3	6.50	7.10
4	2.60	3.85
5	1.79	2.10
Total	3.66	4.35

(ANOVA tests indicate that Year is not significantly related to number of spelling errors)

TABLE 11 – Narrative task: mean original level and mean re-mark level (2002 scripts, typed)

YEAR		Original level	Re-mark level
2002	Mean	4.45	3.87
	N	60	60
	Std. Deviation	.88	.80

TABLE 12 – Discursive task: Mean original level to mean re-mark level by year

Original level attributed	YEAR		Mean original level	Mean re-mark level
3	1995	Mean	3.45	3.70
		N	20	20
		Std. Deviation	.29	.30
	2002	Mean	3.45	3.71
		N	20	20
		Std. Deviation	.29	.53
	Total	Mean	3.45	3.70
		N	40	40
		Std. Deviation	.29	.43
4	1995	Mean	4.47	4.08
		N	20	20
		Std. Deviation	.30	.35
	2002	Mean	4.45	4.13
		N	20	20
		Std. Deviation	.29	.57
	Total	Mean	4.46	4.10
		N	40	40
		Std. Deviation	.30	.47
5	1995	Mean	5.25	4.45
		N	12	12
		Std. Deviation	.19	.36
	2002	Mean	5.23	4.94
		N	12	12
		Std. Deviation	.20	.35
	Total	Mean	5.24	4.69
		N	24	24
		Std. Deviation	.19	.43
Total	1995	Mean	4.26	4.02
		N	52	52
		Std. Deviation	.76	.44
	2002	Mean	4.24	4.15
		N	52	52
		Std. Deviation	.75	.69
	Total	Mean	4.25	4.09
		N	104	104
		Std. Deviation	.75	.58

TABLE 13 – Narrative task: Mean original level to mean re-mark level by year (typed scripts)

Original level attributed	YEAR		Mean original level	Mean re-mark level
3	1995	Mean	3.45	3.31
		N	20	18
		Std. Deviation	.29	.43
	2002	Mean	3.45	3.09
		N	20	20
		Std. Deviation	.29	.40
	Total	Mean	3.45	3.19
		N	40	38
		Std. Deviation	.29	.43
4	1995	Mean	4.45	3.91
		N	20	20
		Std. Deviation	.29	.35
	2002	Mean	4.45	3.82
		N	20	20
		Std. Deviation	.29	.31
	Total	Mean	4.45	3.86
		N	40	40
		Std. Deviation	.29	.33
5	1995	Mean	5.44	4.22
		N	19	19
		Std. Deviation	.29	.52
	2002	Mean	5.46	4.72
		N	20	20
		Std. Deviation	.29	.58
	Total	Mean	5.45	4.47
		N	39	39
		Std. Deviation	.29	.60
Total	1995	Mean	4.43	3.82
		N	59	57
		Std. Deviation	.87	.57
	2002	Mean	4.45	3.87
		N	60	60
		Std. Deviation	.88	.80
	Total	Mean	4.44	3.85
		N	119	117
		Std. Deviation	.87	.70