

Comparability Study of Pupils' Writing from Different Key Stages (2002)

Sylvia Green, Alastair Pollitt, Martin Johnson and Pauline Sutton

A paper presented at the British Educational Research Association Conference
University of Edinburgh, 11-13 September 2003

Contact details

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

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.



UNIVERSITY of CAMBRIDGE
Local Examinations Syndicate

Introduction

This paper reports on the replication of a cross key stage writing study which was carried out in 1995. Cross key stage comparability has become an increasing cause for concern with the high stakes nature of national testing and value added analyses. The aim of the original design of the national curriculum as proposed by the Task Group on Assessment and Testing (TGAT, 1988) envisaged a common scale with levels consistent across key stages. However, it could be argued that levels cannot be equivalent because of curriculum differences and the different experiences and maturity of children at different key stages.

Ideally standards in writing should be consistent both across key stages and between years. Cross key stage comparability depends on: the comparability of writing prompts or activities; the use of a single set of criteria for assessment; the systematic application of those criteria by different assessors. The use of a single set of criteria, as envisaged in the TGAT report, led to the development of level descriptions that form the framework for national curriculum assessment.

Background

The principle of cross key stage comparability rests on the premise that an age independent scale of progression exists. This was a main feature in the assessment system proposed by TGAT. Problems were subsequently recognised and in 1993 the Dearing review addressed some of the issues which had been raised. Although the Dearing proposals addressed a range of problems, it could be argued that the recommendations exacerbated the difficulties for cross key stage comparability. The aim of the ten level scale was to promote the concept of progression by introducing a system of criterion referencing that would lead to comparability between key stages to ensure that the requirements of a particular level should not depend on the age of the student. Black (1998) points out that the opponents of the ten level scale 'appeared to

have nothing to say about the TGAT arguments on the importance of progression, on assessment as a guide to learning, and on continuity between key stages.'

One of the advantages of the age independent scale proposed by TGAT was that there would be a common set of benchmarks which would apply at key stages 1, 2 and 3. The alternative was to have a different scale at each key stage providing age specific levels of attainment. One problem with age specific levels is that a pupil could make real progress but their reported grade could be the same or lower, despite steady progress having been made. This was a key factor taken into account in the TGAT report since this type of feedback has been shown to have a negative impact on learners. Dweck (1986) has shown that a situation where learners achieve the same level at ages 7, 11 and 14 can reinforce the notion that ability is fixed and this can have a negative effect on motivation and self esteem.

The need to report in a way that allows for progression was highlighted in the Cockcroft Report (1982) after which graded mathematics assessment schemes were developed and these developments influenced the TGAT proposals,

The need for differentiation at any particular age, coupled with the overlap of achievements between reporting ages, supports our proposal for a single sequence of levels across the age range for national assessments....Only one set of criteria is required.

(TGAT Report, 1988, X1.109)

William comments that,

The ten level framework was the result of a clear priority to provide a system that allowed students to experience progression (in order to promote a view of attainment as incremental rather than as fixed) and that ensured the focus was on progress, rather than upon absolute levels of achievement. (William, 2001, p.7)

The TGAT model is based on the assumption that it is possible to define progression in a given subject. It could be argued that for some subjects e.g. science and

mathematics, progress is more easily defined, whereas, for English it is more problematic. In the early stages of national assessment, detailed and prescriptive criteria were used to define performance at a given level. However, as this more rigorous method was relaxed and replaced by a system which required more holistic judgements based on generic descriptions, the definition of progress became more problematic.

The development of more holistic level descriptions, which needed to be interpreted alongside the programmes of study, created problems because when criteria for assessment require interpretation, reliability decreases. As William (1993) argues, 'no criterion, no matter how precisely phrased, admits of an unambiguous interpretation'. He goes on to state that for a criterion to be useful for distinguishing levels of performance 'we have to use norms, however implicitly, in determining the appropriate interpretations', and therefore, 'the criterion is interpreted with respect to the target population'. As Angoff (1974) commented, 'one only has to scratch the surface of any criterion-referenced assessment system in order to find a norm-referenced set of assumptions lying underneath'. If age independent scales are to function in a meaningful way William argues that,

The interaction between key stage, based on a student's age, and level, based on his or her attainment, is of crucial importance.

Ultimately, the comparability of levels attained in different key stages depends on the compatibility of the programmes of study.

(William, 1996, p.137)

Any lack of consistency in the definitions of progress from one key stage to another can cause problems for those using the language of 'levels'. This is the case when information about pupil performance is transferred from one teacher to another and / or from one key stage to another. As Pollitt comments,

As children pass from one key stage to the next it would seem essential that their attainment keeps its value, that is, that a level 5 in key stage 3 should mean the same standard as level 5 at key stage 2. (Pollitt, 1994, p.67)

The TGAT report recommended that there was a need for continuity in assessment, and in curriculum planning, from age 5 to 16, with progression emphasised across the transition from primary to secondary work. The recommendations on transfer information emphasised that the data from primary schools should make clear what pupils know, understand and can do at the end of the primary phase in relation to the attainment targets so that continuity and progression are more likely to be achieved. The danger is that the levels at each key stage are, in reality, age dependant and that decisions made about levels of performance are based on an inherent system of norm referencing within the key stage rather than progression through a ten level scale.

The TGAT report recognised the importance of moderation as an essential part of an assessment system,

We are all, as individuals, persuaded that those things which occur frequently in our experience are 'normal'....In the absence of equally powerful external evidence, teachers' expectations become the teachers' standards.... In the absence of a close definition of what to look for and how to observe it, we look for confirmation of our expectations.

(TGAT Report, 1988, X.65)

Sainsbury and Sizmur (1998) suggested that in order to discern unity in the attainment targets of the national curriculum, we need to look outside the words of the level descriptions themselves to a pre-existing understanding which defines the necessary yardstick. As Wiliam (1993) points out, 'there then develops a case law of curriculum designers and test developers'. To those groups could be added 'teachers', since their case law is developed at the point of teaching, learning and assessment.

The risk is that if assessment criteria become more abstract and complex, requiring greater interpretation, then the assessment can become less objective. Sizmur and Sainsbury argue that this need not be the case,

The better we understand the nature of the subject, how it is taught, how pupils learn and make progress in it, the better we are able to rate specific pieces of work in terms of quality....Level descriptions are in one sense a means of imposing coherence on diverse elements of attainment. (Sizmur and Sainsbury, 1997, p.11)

They advise that teachers need to get to grips with the well-rounded attainment scale that lies behind the level descriptions, to understand it thoroughly and to know how to apply it in their assessments and to understand levels of performance and the nature of progression. Hall and Harding (2002) support this view, 'teachers need to interpret the loosely framed level descriptions through a well-defined community of practice'.

There are implications for the development of tasks and tests, as Sainsbury and Sizmur (1998) point out. National curriculum tasks and tests need to be based on the programmes of study but calibrated to the level descriptions while reflecting their abstract and complex nature and as Pollitt warned,

We are in danger of implementing a system of tests that behave like thermometers, all pretending to measure on the Celsius scale, but which actually each have their own freezing point and each their own idea of what constitutes a nice summer's day.

(Pollitt, 1994, p.69)

Aims of the study

To investigate:

(i) *equivalence of standards of judgement at different key stages;*

Markers' judgements were analysed to find out how close they were as they judged pupils' level 3 performance at key stages 1 and 2, and level 5 performance at key stages 2 and 3. Markers commented on the salient features in their decision making and their reports were used to determine on which components of writing ability different groups of markers (by key stage experience) expected the same absolute standards and on which they tolerated differences.

(ii) *underlying features of writing compared across key stages;*

A sub-sample of scripts was analysed to consider the underlying features of writing judged to be at the same level and to compare them across key stages. The aim of the comparisons was to find out what key stages 1 and 2, level 3 performances had in common and what key stages 2 and 3, level 5 performances had in common.

(iii) *changes over time from 1995 to 2002.*

Implications for comparability of standards over time were considered in the light of changes from 1995 to 2002.

Method

The Sample

The experimental study used a common task (see Appendix 1, p.17) to provide empirical evidence on the comparison of pupils' writing at the same level in different key stages. The main focus was on two comparisons, between samples of children deemed to be at level 3 in key stages 1 and 2 or at level 5 in key stages 2 and 3. All children in the classes selected were asked to participate as it would have been difficult for teachers to include only the children at the focus levels (level 3 and level 5).

The schools which took part in the 1995 study were invited to participate in the replication and they all agreed. It was important to involve the same schools in the replication study so that comparisons could be made over time. They included a range of school types and sizes with each school providing 25 – 30 pupils. 15 primary schools and 9 secondary schools were involved. Teachers were asked to include all children in their classes, apart from those for whom they felt it would be wholly inappropriate e.g. those with special needs statements for writing. For the secondary schools, where there were ability sets, the middle sets were selected since the target level for key stage 3 was level 5 and this was a common level in the middle ability groups. In primary classes there were mixed ability groups and so the whole range was included. In order to maintain similar demographic characteristics and to minimise the management problems for teachers the final sample for the 2002 study was bigger than was necessary to achieve the target of 50 pupils per level at each key stage. The 2002 sample included

232 pupils at key stage 3, 539 pupils at key stage 2 and 478 pupils at key stage 1, 1249 pupils in all (see Appendix 2, p.18).

The Task

Four factors combine to make it difficult to equate standards set between different key stages.

- Different tests set different demands and offer different opportunities.
- Task differences may not matter if they are not reflected in the mark schemes, while mark scheme differences could lead to different outcomes for the same performance.
- Markers from different backgrounds, for example primary and secondary, are likely to interpret mark schemes differently and marking may also be affected by the distributions of performance that markers are used to seeing. Level 3 is a 'below average' performance at key stage 2, but 'above average' at key stage 1.
- Different tasks produce different performances which may be evaluated differently. Exemplars used to train markers may set different standards.

In order to compare the performances of pupils deemed to be at the same level but from different key stages, it was necessary to address these problems. It was therefore decided to use a common task across key stages, with identical mark schemes and common markers, calibrated on the same sample material. This allowed direct equating of performances from different key stages. The narrative task chosen was similar in style to one used in the 1995 key stage 2 national reading test (see Appendix 1, p.17).

Marking

A common mark scheme was used based on the level descriptions for writing from the National Curriculum (see Appendix 3, p.19). The markers were asked to take a holistic view when matching a response to the performance criteria, using the best-fit principle. They were instructed not to treat any aspect of the criteria as a hurdle for achieving a particular level. Having assigned a level, they were then asked to distinguish within a level using a 10 point scale (i.e. a low level 3 would be 3.1 or 3.2 and a high level 3 would be a 3.8 or 3.9).

Markers were selected from key stages 1, 2 and 3 so that judgements of markers from different backgrounds could be compared. They came together for a co-ordination meeting after they had marked co-ordination scripts. Exemplar scripts, with commentaries for each level, were used to inform judgements and during co-ordination markers had the opportunity to discuss their judgements and to agree standards in the light of the exemplar materials.

The scripts from the three key stages were mixed and distributed to the team of markers so that each marker marked scripts from across the full age range. Each script was marked twice and the first marker in each case was instructed not to write on the script or to indicate the level they had given. On a separate sheet of paper they gave a brief explanation of the level given. These comments were used later to inform the analyses of how judgements were made by markers from different key stages at the target levels (level 3 / level 5).

Data

For each pupil the following data were collected: school; date of birth; gender; key stage; teacher assessment level for writing; English / writing national test score; levels awarded by two markers; final level awarded; markers' identities.

Evaluation of Marking

Statistical comparisons were carried out to find out how close markers' judgements were when they assessed performance at level 3 in key stages 1 and 2, and at level 5 in key stages 2 and 3. The equivalence of standards of judgement at different key stages was further investigated using markers' reports in which each marker identified features in pupils' writing that were most and least significant when assigning a level 3 or a level 5 performance.

Descriptive Comparisons

A member of the project team analysed a sub-sample of scripts, described the underlying features of the writing and compared them across key stages. Scripts considered 'typical' for their level and key stage were identified by relevant members of the marking team for each key stage. Those scripts awarded the same levels (level 3 and level 5) were selected from the appropriate key stages i.e. level 3 scripts from key

stages 1 and 2, and level 5 scripts from key stages 2 and 3. Where possible the comparisons involved writing within the same range of a level.

The questions considered were:

What do key stage 1 and key stage 2, level 3 performances have in common?

What do key stage 2 and key stage 3, level 5 performances have in common?

The aim was to consider differences and similarities in linguistic features of the writing as well as maturational differences in performance between the key stages.

Findings and discussion

Key Stages 1 and 2 comparability at level 3

There was very close agreement between the Gate levels awarded to children from key stages 2 and 3 who achieved level 3 in the national tests, (see Table 1).

Table 1 Gate levels for children with level 3 in national tests

	KS1	KS2
Gate level	3.17	3.12

Level 3 is represented by 3.5 and so these Gate levels indicate that the key stage 1 and 2 children did not perform as well in the Gate task as their national test levels would suggest. Part of this difference could be attributed to the fact that the national tests for English include reading and reading levels are higher than writing levels. Another factor is motivation since we would expect national test levels to be higher because of test practice and the high stakes nature of national testing. Also, we need to take into account the fact that the Gate writing was completed two months before the national tests and therefore there is a difference in development that will have affected writing performance.

It is interesting to note that the gate level for key stage 1 is very slightly higher than for key stage 2. The analyses of matched level 3 scripts showed that children tried to concentrate on a wider range of punctuation and sentence structure at key stage 2, often unsuccessfully. In doing so there was evidence that they did less well in the more basic skills such as sentence demarcation, with more comma splice errors and omitted

punctuation. This may highlight the importance of having time to consolidate and re-visit skills learned earlier as new features of language are introduced. There were, however, aspects of writing where the key stage 2 performances were better e.g. reader awareness and style. These are qualitative features and can be difficult to 'pin down' and therefore may not have carried as much weight in the 'best-fit' process.

There was also very close agreement between Gate levels awarded to children with level 3 teacher assessments for writing (TAs), (see Table 2).

Table 2 Gate levels for children with TA level 3

	KS1	KS2
Gate level	3.27	3.38

The TA levels match the Gate levels more closely than the national test levels did as might be expected since both assess writing performance rather than reading and writing as in the national tests. The fact that the key stage 1 Gate levels are lower than those at key stage 2 may result from a 'halo effect' with key stage one teachers being more generous in their judgements since their level 3 writers would be the 'best of the bunch'.

The evidence of close comparability at level 3 between key stages 1 and 2 is supported by the evaluation of marker judgements. There was close agreement between markers about what constituted a level 3 piece of writing. Text level features such as structure and organisation were most important with the emphasis on the overall shape, coherence, pace and control of the writing. Technical aspects were considered less important although sentence demarcation was more important where its omission compromised clarity and readability.

Key Stages 2 and 3 comparability at levels 4 and 5

The evidence led to a more complicated picture with less agreement between key stages 2 and 3 (see Table 3).

Table 3 Gate levels for children with English national test levels 4 and 5

National test level	Gate level KS2	Gate level KS3
L4	3.77	4.51
L5	4.76	5.56

KS2 children who scored levels 4 or 5 in national tests achieved lower Gate levels than their KS3 counterparts, by a margin of about $\frac{3}{4}$ of a level. The key stage 2 Gate levels were lower than the national test levels suggested they should be, also by $\frac{3}{4}$ of a level. There are some mitigating factors that may account for some of these differences. As in key stage 1, we might expect the Gate levels to be lower than the national test levels because of the 'reading' effect that would increase the national test levels. Whereas, the Gate task was an assessment of writing and so the levels would be lower. Also motivation and age difference would also lead to lower Gate levels and these factors may account for some of the difference at key stage 2. However, the KS3 Gate levels at levels 4 and 5 match the national test levels almost perfectly and so we are left with the question – why do these factors not have the same effect at Key stage 3?

A number of explanations could account for some of this cross key stage difference. It may be that the 'gap' between reading and writing narrows at key stage 3 as the focus and demand of the curriculum change, perhaps with more emphasis on different kinds of writing and reading activities. Part of the explanation is likely to be that the key stage 3 reading test is more a test of writing than is the case at key stage 2, with long written responses. We might therefore expect the Gate task and the key stage 3 English test to produce more similar results. One other possibility is that the standards set in the key stage 2 English test are more lenient than those set by the key stage 3 national test. Evidence from the 1995 study suggested that there was greater cross key stage comparability between key stages 2 and 3 than the replication has indicated. This could mean that there have been changes in standards during the intervening years. In a study to compare test standards over time (Massey et al, 2002), there was evidence to

suggest that test standards for key stage 2 English had slipped from 1996 to 2001 and that much of this was due to changes in the reading element of the tests.

The evaluation of marking at level 5 indicated that there was some agreement between markers about what constituted a level 5 piece of writing. As at level 3, text level features were more important than technical aspects of the writing. Structure, control, pace, coherence and style were significant features as well as descriptive language, genre awareness and audience. However there were indications that key stage 3 markers expected more for the level and this could mean that it is more difficult to achieve level 5 at key stage 3.

The comparisons of matched scripts showed that level 5 writing at key stage 3 was better in a range of ways than at key stage 2. KS3 writing was more sophisticated, with more mature language and style. However, these differences in the quality of the writing were not reflected in the levels awarded, perhaps because they were more difficult to define. This may also be because the level criteria did not enable the markers to 'credit' the improved quality of the writing and this may have led to a 'ceiling' effect. The new mark schemes introduced in 2003 are more detailed with writing strands and mark bands within each of these. This more detailed marking may enable markers to discriminate more effectively between different pieces of writing, however, there is a danger that such 'deconstruction' of writing into finer criteria may lead to reduced reliability. Decisions will need to be made on smaller units of learning with marks awarded and then reconstructed to give an overall mark that has then to be converted to a national curriculum level. Evaluations of the new mark schemes will provide insights into the merits and demerits of the new system.

There were differences in cross key stage comparability for teacher assessment (TA) at key stages 2 and 3. However, the evidence showed a reversed pattern, with the key stage 3 TA appearing to be more lenient than at key stage 2, (see Table 4).

Table 4 Gate levels for children with TA levels 4 and 5

TA level	Gate level KS2	Gate level KS3
L4	4.33	3.90
L5	5.25	4.82

The key stage 3 Gate levels were lower than those at key stage 2 according to those suggested by the TA levels, indicating that the key stage 3 teachers had been overgenerous by a margin of about half a level. The TA levels matched the Gate levels more closely at key stage 2 than the national test levels had done. This would be expected since both the Gate task and the teacher assessments assessed writing performance. However, we are left with the question – why were the key stage 3 teacher judgements so generous?

The matched script comparisons indicated that level 5 writing was better at key stage 3 than at key stage 2, even though the same level was given. It could be argued that the teachers were not overgenerous but were in fact more realistic. It is also possible that the teachers were making judgements based on a range of different types of writing that were more representative of the key stage 3 curriculum. They may have felt that their students had improved during the key stage across a wider range of writing and that in doing so narrative may have been less important than it had been at key stage 2. Perhaps key stage 3 writers felt that the Gate task was ‘immature’ and therefore did not fully engage with it. So the fact that their performance in the Gate task was poorer than suggested by their TA levels may have resulted from a combination of factors related to maturity, the focus of the curriculum for writing and the extent to which the Gate task could represent the development of KS3 writing abilities.

Whatever the reasons, and taking the various mitigating factors into account, there is still a sizeable difference between key stages 2 and 3. The possible reasons for cross key stage differences raise a number of interesting questions, made more interesting because of the pattern of differences between national test and teacher assessment comparability. Any explanation must lie in a combination of areas including: test difficulty and standard setting; the extent to which the marking criteria encapsulate the development of narrative writing; and the extent to which teacher assessment and national tests address the same constructs at different key stages. Perhaps we should consider the nature of the traits that we are assessing and the way they change and develop from key stage 1 to key stage 3. If the ‘gap’ between reading and writing does narrow as children move through the key stages, is it because they improve in writing

and 'catch up', is it because our measuring instruments have changed or perhaps we mean something different when we talk about 'reading' and 'writing'?

These issues have implications at a time of high stakes testing and value added analyses. If level 5 writing at key stage 2 means something different from level 5 writing at key stage 3, then there are problems for those who use such information in their decision-making. *The findings from this study illustrate the fact that assessment is difficult and complicated even when we think we are addressing 'straightforward' subjects like reading, writing or English. It could be argued that it would be wiser to concentrate on teacher assessment and to accept that there will be a price to pay in lower reliability. The reward would be that pupil level information could be used to support diagnostic and formative assessment and to help teachers to help children to learn.*

References

Angoff, W.H. (1974) Criterion referencing, norm referencing and the SAT. *College Board Review*, 92, pp. 2-5.

Black, P. (1998) Learning, League Tables and National Assessment: opportunity lost or hope deferred? *Oxford Review of Education*, Vol.24, No.1. pp. 57-68.

Cockroft Report (1982) *Mathematics Counts: Report of the Committee of Inquiry into the Teaching of Mathematics in Schools*. London: HMSO.

Dearing, R (1993) *The National Curriculum and its Assessment: interim report*. London: National Curriculum Council and Schools Examinations and Assessment Council.

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

Dept for Education and Employment (1999) *English: The National Curriculum for England, Key stages 1-4*. London: DfEE/QCA

Dweck, C.S. (1986) Motivational Processes Affecting Learning. *American Psychologist* (Special Issue: *Psychological science and education*), 41(10), pp. 1040-1048.

Hall, K. and Harding, A. (2002) Level Descriptions and Teacher Assessment in England: towards a community of assessment practice. *Educational Research*, 43, 4.

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

National Curriculum Task Group on Assessment and Testing (TGAT) (1988) *A Report*. London: Department of Education and Science.

Pollitt, A. (1994) 'Measuring and Evaluating Reliability in National Curriculum Assessments' in Hutchinson, D. and Schagen, I. (eds) *How Reliable is National Curriculum Assessment?* London: NFER.

Sainsbury, M. and Sizmur, S. (1998) Level Descriptions in the National Curriculum: What kind of criterion-referencing is this? *Oxford Review of Education*, Vol.24, No.2, pp. 181-193.

Sizmur, S. and Sainsbury, M. (1997) Criterion referencing and level descriptions in National Curriculum assessment. *British Journal of Curriculum and Assessment*, Vol.7, No.1, pp. 9-11.

William, D. (1993) Validity, dependability and reliability in National Curriculum assessment. *The Curriculum Journal*, Vol.4, No.3, pp. 335-350.

William, D. (1996) National Curriculum Assessments and Programmes of Study: validity and impact. *British Educational Research Journal*, Vol.22, No.1, pp. 129-141.

William, D. (2001) *Level Best? Levels of attainment in National Curriculum assessment: a report*. London: The Association of Teachers and Lecturers.

"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.



Appendix 2

Using national results for 2000 we estimated the following approximate percentages of each cohort at each level.

KS	1	2	3	4	5	6	7	8
	below L4		11	21	35	21	6	1
2	below L3		6	17	46	29		
1	11	31	20					

The target comparisons are shown in boxes. The number of children needed in the total sample was calculated (assuming that the samples are nationally representative) based on the assumption that each sub-sample should contain at least 50 scripts.

KS	1	2	3	4	5	6	7	8	
3	below L4		16	30	50	30	9	2	137
2	below L3		18	50	135	85			288
1	28	78	50						156

A minimum sample of 137 pupils at key stage 3, 288 at key stage 2 and 156 at key stage 1 would therefore need to be identified.

Appendix 3 Performance criteria for writing

Level	Performance criteria for writing
1	The writing communicates meaning through simple words and phrases. The pupil may show some awareness of how full stops are used. Letters are usually clearly shaped and correctly oriented.
2	The writing communicates meaning using some features of a narrative or non-narrative form appropriately. The choice of vocabulary is apt and interesting showing some awareness of the reader. Ideas are developed in a sequence of sentences, sometimes demarcated by capital letters and full stops. Simple, monosyllabic words are usually spelt correctly, and where there are inaccuracies the alternative is phonetically plausible. In handwriting, letters are accurately formed and consistent in size.
3	The writing shows evidence of organisation, imagination and clarity. The main features of the chosen form are used appropriately, beginning to be adapted for a reader. Sequences of sentences extend ideas logically and words are chosen for variety and interest. The basic grammatical structure of sentences is usually correct. Spelling is usually accurate, including that of common, polysyllabic words. Punctuation to mark sentences – full stops, capital letters and question marks – is used accurately. Handwriting is joined and legible.
4	The writing is lively and thoughtful with ideas sustained and developed in an interesting way and organised appropriately for the purpose and the reader. Vocabulary choices are often adventurous and words are used for effect. The pupil is beginning to use grammatically complex sentences, extending meaning. Spelling, including that of polysyllabic words that conform to regular patterns, is generally accurate. Full stops, capital letters and question marks are used correctly, and pupils are beginning to use punctuation within the sentence. Handwriting style is fluent, joined and legible.
5	The writing is interesting, conveying meaning clearly in the chosen form for an intended reader. A more formal style used where appropriate. Vocabulary choices are imaginative and words are used precisely. Simple and complex sentences are organised into paragraphs. Words with complex regular patterns are usually spelt correctly. A range of punctuation, including commas, apostrophes and inverted commas, is usually used accurately. Handwriting is joined, clear and fluent.
6	The writing generally engages and sustains the reader's interest, showing some adaptation of style and register to the chosen form, such as the use of an impersonal style where appropriate. The pupil uses a range of sentence structures and varied vocabulary to create effects. Spelling is generally accurate, including that of irregular words. Handwriting is neat and legible. A range of punctuation is usually used correctly to clarify meaning, and ideas are organised into paragraphs.
7	The writing is confident and shows an appropriate choice of style in the chosen form. In narrative writing, characters and settings are developed and, in non-fiction, ideas are organised and coherent. Grammatical features and vocabulary are accurately and effectively used. Spelling is correct, including that of complex irregular words. Work is legibly and appropriately presented. Paragraphing and correct punctuation are used to make the sequence of events or ideas coherent and clear to the reader.
8	Pupils' writing shows the selection of specific features or expressions to convey particular effects and to interest the reader. Narrative writing shows control of characters, events and settings, and shows variety in structure. Non-fiction writing is coherent and gives clear points of view. The use of vocabulary and grammar enables fine distinctions to be made or emphasis achieved. Writing shows a clear grasp of the use of punctuation and paragraphing.