

# **Guidance on the Principles of Language Accessibility in National Curriculum Assessments**

Research Background



Dahlia Janan and David Wray

September 2012

Ofqual/12/5217

# Contents

Introduction.....	2
Assessment questions: context and content.....	3
The nature of readability.....	6
Readability: looking at text features.....	9
Linguistic features.....	11
Word difficulty.....	11
Sentence difficulty.....	15
Cohesion and coherence.....	22
Content structure and complexity.....	28
Legibility and print issues.....	31
Text organisation.....	35
Readability: looking at the characteristics of readers.....	36
Physical capabilities.....	39
Reading abilities.....	39
Engagement/motivation.....	40
Prior knowledge.....	41
Gender.....	43
Accommodations, modifications and universal design.....	44
Conclusion.....	48
References.....	49
Appendix: The 1,000 most commonly used words in English text.....	58

## **Introduction**

This review accompanies the document, which describes the principles which should guide the development of clear assessment questions. The purpose of the review is to present and discuss in detail the research underpinning these principles. It begins from the standpoint that National Curriculum assessments, indeed any assessments, should be:

- appropriate to the age of the pupils
- an effective measure of their abilities, skills and concept development
- fair to all irrespective of gender, language, religion, ethnic or social origin or disability. (Ofqual, 2011)

The *Regulatory Framework for National Assessments: National Curriculum and Early Years Foundation Stage* (Ofqual, 2011) sets out a number of common criteria which apply to all aspects of the development and implementation of National Assessments. One of these criteria refers to the need for assessment procedures to minimise bias: “The assessment should minimise bias, differentiating only on the basis of each learner’s ability to meet National Curriculum requirements” (Section 5.39, page 16). The Framework goes on to argue that: “Minimising bias is about ensuring that an assessment does not produce unreasonably adverse outcomes for particular groups of learners” (Annex 1, page 29). This criterion reinforces the guiding principle that any form of assessment should provide information about the knowledge and understanding of relevant content material. That is to say that the means through which this knowledge and understanding is examined, the design of the assessment and the language used should as far as possible be transparent, and should not influence adversely the performance of those being assessed.

There is clearly a large number of ways in which any given assessment task can be presented and in which questions can be asked. Some of these ways will make the task more accessible – that is, easier to complete successfully – and some will get in the way of successful completion. Section 26 of the *Fair Access by Design* (Ofqual, 2010) document lists a number of guiding principles for improving the accessibility of assessment questions, although the research basis for these principles is not made completely clear in that document. The aim of the current review is to examine the research background more closely in order to provide a more substantial basis for a renewed set of principles to underpin the concept of language accessibility.

In the review, each section will be prefaced by a statement of the principles outlined in *Guidance on the Principles of Language Accessibility in National Curriculum Assessments* and then the research evidence underpinning these principles will be reviewed.

## **Assessment questions: context and content**

### **Principles outlined in the guidance**

- The success of pupils in assessments can be influenced by:
  - the context provided for assessment questions
  - the language in which assessment questions are expressed.
- Readability of assessment questions is a key consideration for assessment designers.

Crisp (2011) has argued that while part of the difficulty of an assessment task will be due to the intrinsic demands of the subject content of that task, the actual difficulty can be affected, sometimes in unexpected or unfair ways, by features of the way that questions are asked (Pollitt and others, 1985; Fisher-Hoch and other, 1997; Ahmed and Pollitt, 2007). Ahmed and Pollitt (2007) argue that:

Putting questions into context inevitably involves using extra words to ask the question. If pupils have to read more text in order to answer a question then their reading ability is being tested as well as their understanding of concepts. (p. 203)

This may be the aim in some assessments, where reading skill is itself being directly assessed. But for assessments in other subject areas, complexities in the way assessment questions are expressed can interfere with the accurate operation of the assessment. Crisp and Sweiry (2006) have also shown that the visual context within which an assessment task is embedded (accompanying pictures and so on) can affect pupils' responses, both in terms of their success with the subject content of a question, but also in affecting their motivation to tackle the task. Crisp and Sweiry quote one pupil who made this point:

The use of pictures isn't particularly useful in trying to answer the question, but it's quite daunting on the day if all you've got is text and you've just got to read it, so maybe a picture would calm your nerves. (p. 146)

Research into the effects on pupils of the contextual variables of assessment questions has a substantial history. Pollitt and other (1985) identified two categories of potential difficulty in any assessment item:

- Concept difficulty – the intrinsic difficulty of the concept itself

- Question difficulty – which may be rooted in the language of the questions, the presentation of questions and the use of mark schemes in rewarding responses.

Other related literature has alerted us to a variety of influences on assessment task difficulty, particularly in the areas of mathematics and science, including context and language.

The Assessment of Performance Unit (1985) claimed that, “Mathematics performance cannot be assessed on its own; the mathematics must be communicated in some way and presentation influences performance” (p. 838). They went on to claim that context – that is the material surrounding a mathematics assessment task such as accompanying pictures and/or the embedding of the task in a real-life situation – could affect the success rate on that task from a few percentage points up to 20 per cent. They showed that the performance of lower ability candidates was improved by the presence of a degree of context, but that a very rich context could also reduce performance. Nickson and Green (1996) later found that the degree of context in which a mathematical question was set could affect pupils’ choice of the correct mathematical operator with which to answer the question.

Shuard and Rothery (1984), in a seminal study of school mathematics language, distinguished three broad categories of mathematical words:

- Words which are specific to mathematics and not usually encountered in everyday language (for example *hypotenuse, coefficient*)
- Words which occur in mathematics and in everyday English, but have different meanings in these two contexts (for example *difference, volume, mean*).
- Words which have the same or roughly the same meaning in both contexts (for example *fewer, between*).

It is in the second of these categories that there is the greatest potential for pupils to be confused by the language of mathematical assessment. Assessment developers clearly need to think very carefully about the words they use in assessment questions, especially with younger age groups in mind. Assessment of mathematics should, after all, be just that and not assessment of pupils’ linguistic skills or their abilities to ‘think like the assessor’.

Schagen and Sainsbury (1996) have confirmed that reading ability can make a significant contribution to pupils’ scores on National Curriculum mathematics assessments, and the same conclusion can be drawn from the study by Shorrocks-Taylor and other (2003) of the effects of different question types on pupil performance in the Key Stage 2 National Curriculum Mathematics assessment. This

study found that the substitution of what the authors refer to as contextual number questions (problem solving) by number-focused data-handling questions improved the assessment scores of a number of the pupils taking this assessment. The suggestion is that the embedding of number questions within a heavily language-reliant context had made certain questions more difficult for certain (but not all) pupils to answer successfully.

A similar picture emerges from research into the effects of language on learning, and hence assessment of learning, in science. Fang (2006), for example, has investigated the linguistic demands of school science texts and concluded that these can make a significant difference to pupil understanding of these texts. Fang's work supports the observation of Wellington and Osborne (2001) that for many pupils the greatest barrier to learning science lies in learning its language. If this is the case then it is likely that this language may remain a barrier to pupils performing their best in assessments of their science knowledge and understanding, with this assessment often, necessarily, being carried out through the medium of language.

Therefore, the accessibility of the language through which assessments are made is a crucially important consideration for the designers of these assessment instruments. The language used needs to be readable in the broadest sense, and the principles underpinning success are the principles underpinning the concept of readability. A number of factors influence the readability of any text, and test designers need to be alert to the influence of these factors.

## The nature of readability

### Principles outlined in the guidance

- Readability is influenced by:
  - characteristics of the text itself
  - **and** characteristics of the readers of the text.
- Writers of assessment questions need to take both of these dimensions into consideration.

Readability is the study of matching a reader and a text (Gilliland, 1975). It has been argued that the most important pedagogic decision that teachers make is “making the match” (Fry, 1977); that is, ensuring that pupils are supplied with reading materials, in whatever subject, that are at an appropriate level of difficulty for them. Pupils who are given reading materials that are too easy are not sufficiently challenged and their learning growth can be stunted (Chall and Conard, 1991). Pupils who are given reading materials that are too difficult can fail to make progress (Gambrell and other, 1981), are frequently off task and may exhibit behavioural problems (Anderson and other, 1987), or may become so frustrated that they simply give up (Kletzien, 1991). Making the match is therefore a crucial skill for teachers, and it has long been considered that the successful exercise of this skill requires knowledge of the readability level of materials. The Bullock Report commented that:

a particularly important teaching skill is that of assessing the level of difficulty of books by applying measures of readability. The teacher who can do this is in a better position to match children to reading materials that answer their needs. (DES, 1975, p. 113)

Similarly, without an assessment of the readability of assessment questions, the test developer risks producing items that do not correctly match the reading abilities of the pupils for whom the assessment is planned. If the readability level of a test item is higher than the reading ability of the pupil then it is likely that the item is not assessing the construct of interest (the subject matter) but rather the pupil’s reading ability.

Defining and exploring this concept gave rise to a significant body of research from the 1920s to the early 1990s. One of the major outcomes of the research was the production of a large number of ‘readability formulae’; that is, approaches to analysing texts which were designed to give a quantitative measure of the ‘level’ a

reader would need to be at in order to read and understand a particular text successfully. Various definitions of the concept of readability have emphasised:

- the elements in a text which are associated with comprehension (or lack of it) on the part of the reader: that is, the understanding of words, phrases and ideas in the passage
- a person's ability to read a given text at an optimum speed
- motivational factors which affect a reader's interest in reading a text.

According to Dale and Chall (1948), these three elements of the definition of readability were not separate, but interacted with each other. To explain this interaction, Gilliland (1975) provided the following example:

... in a scientific article, complex technical terms may be necessary to describe certain concepts. A knowledge of the subject will make it easier for a reader to cope with these terms and they, in turn, may help him to sort out his ideas, thus making the text more readable. This interaction between vocabulary and content will affect the extent to which some people can read the text with ease. (p. 13)

Thus, definitions of readability have never been entirely text-centric. However, despite the established claim put forward by Harris and Hodges (1995, p. 203) that "Text and reader variables interact in determining the readability of any piece of material for any individual reader", approaches to the measurement of readability have not generally reflected such interactive definitions. Readability measurement has instead usually involved objective estimates of the difficulty level of reading material derived from the application of formulae which generally took into account sentence and vocabulary difficulty.

Most studies of readability have been carried out within a positivist paradigm (Janan and other, 2010). According to this paradigm, the difficulty of a text was determined by factors within the text itself, and reading was seen as a matter of getting meaning from the page. However, views about the nature of the reading process have changed over the past 20 years or so towards a more interpretive definition which emphasises that making meaning through reading comes from a process in which the reader's mind interacts with the text. As the result of this new interpretive paradigm of reading, research into readability has also needed to change. Hence, it was appropriate that the study of readability should also shift into the interpretive paradigm (Janan and other, 2010). In this review, we have tried to allow for these two dimensions of the readability question by focusing firstly upon factors within the text itself, and secondly upon characteristics of readers.

The questions which we will try to answer in what follows are two-fold:

- What can affect the readability of National Curriculum assessment questions in terms of the features of the texts employed in these items?
- What might affect the readability of National Curriculum assessment questions in terms of the characteristics of the target readers of these items?

## **Readability: looking at text features**

A number of text features affect the readability of texts. These are:

- linguistic features:
  - word difficulty
  - sentence difficulty
- cohesion and coherence
- content structure and complexity
- legibility and print issues
- text organisation.

In the following sections we will review what is known from research into the effects of a variety of text-based features on readers' understanding of texts. The effects of some of these features are universal in operation – font size and style have, for example, been shown to affect reading success in adults as well as in children. The effects of some others are, naturally, influenced by the age and/or reading maturity of the reader. One would expect, for example, that older pupils might be able to take fuller meaning from a wider range of vocabulary and sentence structures than would younger readers.

Matching text features to reader age and maturity is not an exact science, and indeed the research basis for a model of reading development which would offer clear guidance in this area is debatable. Some researchers, building on the work of Chall (1976), have suggested a developmental model of reading, paralleling the developmental cognitive model of Piaget, which posits that readers go through a number of stages of development. Chall's original (1976) model suggested five developmental stages:

- **Stage 1: initial reading or decoding stage.** The reader's central task is the learning of arbitrary letters and the association of these with the corresponding parts of spoken words, largely through the use of letter–sound correspondences (phonics).
- **Stage 2: confirmation, fluency or automaticity stage.** This is essentially a stage of consolidation during which the reader gains more and more control over the processes learned in Stage 1. It is essentially a stage of practice; that is, increasing experience of reading a range of texts.

- **Stage 3: reading for learning.** At this stage, readers develop their ability to bring prior knowledge to their reading and begin to use their reading to acquire facts.
- **Stage 4: multiple viewpoints stage.** Readers develop their abilities to integrate and critique information gained through reading a variety of sources.
- **Stage 5: construction and reconstruction stage.** Readers are able to apply their reading skills to a variety of contexts and situations. They recognise when they do not understand a text and are able to take appropriate action.

Models such as this are, naturally, popular with publishers of reading schemes, and teachers of reading to some extent, because they seem to offer a neat structure to guide text development and the emphasis of teaching. The current eight-level model of reading development in the National Curriculum derives from this approach.

Other researchers have questioned the application of such 'stage' models, largely by doing exactly what critics of Piaget's cognitive development model did: finding lots of examples of pupils who did not perform in the ways the model predicted they should. The key area for these critiques was the nature and practice of reading (and writing) in very young pupils. Clay's (1977) concept of 'emergent literacy' inspired a huge number of research studies during the 1990s whose common thread was that young pupils were quite able to engage in literate behaviours, some which could be interpreted as being at Chall's Stage 3 or 4, before they had technically mastered Stage 1. A major theoretical outcome of this research was the understanding that pupils' performance as readers was not best described by placing them on a 'scale' of reading skill development, but was rather determined by the complex interaction of a number of factors such as prior experience, motivation and context of operation. The impact of some of these factors on readability is discussed in 'Readability: looking at the characteristics of readers', on page 38 of this review.

Therefore it is difficult to be precise about the age and reading maturity level at which particular linguistic difficulties in a test item will have a major negative impact upon test performance.

## Linguistic features

### Word difficulty

#### Principles outlined in the guidance

- Word difficulty can play a part in the readability of assessment questions. But there are some caveats which writers of these questions need to take into consideration:
  - The length of a word is an uncertain guide as to its difficulty, **and**
  - The characteristics of the readers of the text are not the same.
  - The familiarity of a word to its target audience is a more useful indicator, **but...**
- Lists of familiar or frequently occurring words need to be examined carefully to ensure they relate to the target audience.

Word difficulty has to do with the pupil's understanding of individual words. The difficulty of words is measured in two ways:

- the length of the word
- the familiarity of the word.

#### The length of the word

Word difficulty has traditionally been measured by the length of the word, with the assumption that longer words are harder to read than short ones. Word length is measured in two ways – the numbers of letters and/or the number of syllables in the word. It is often suggested that short words are perceived as more familiar and long words as more formal or technical. There is research that shows that readers pause for longer when reading longer words (Just and other, 1982).

Nevertheless, there have also been findings that suggest the failure of the assumption that short words are always easier to read than long ones. There are examples of monosyllabic words (for example *adze*, *gneiss*) found in lower secondary school text-books which are unlikely to be easy words for the pupils who read such books (Perera, 1980). Accordingly, the assumption that short words are easier to read than long words must be viewed with caution.

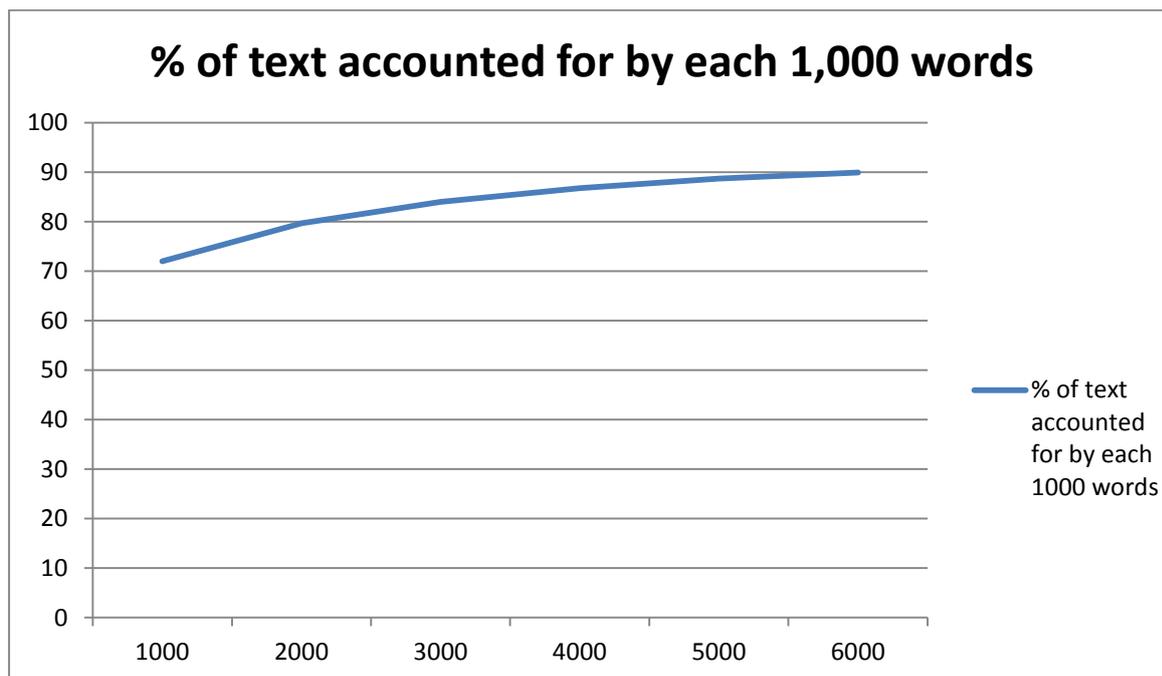
Neither is it always the case that longer words are harder to read. There are very few 7 to 11-year-olds, for example, who will not be able to read and understand words such as *tyrannosaurus* and *diplodocus*. On a recent visit to a class of 5-year-olds, one of us overheard a group of pupils announce that one of them was *tachycardic* (they were engaged in improvised play as doctors and patients and had probably encountered this word while watching a popular television hospital drama). It is unlikely that these pupils could, at that point, read this word, but research (and common sense) suggests that pupils succeed more quickly in learning to read words which are already in their spoken vocabulary.

These examples suggest that the length of a word is not the crucial feature in determining whether it can be read easily or not. Pupils' motivation to read a word and their existing familiarity with it are much more significant indicators of reading ease.

### Word familiarity

Word difficulty is affected by word familiarity. In previous readability research, and in some readability formulae, word familiarity usually refers to those words that appear in word lists such as the Dale-Chall (1948) list (revised in 1995 – see Chall and Dale, 1995 – consisting of 3,000 words). It is presumed that words which appear on this list will be relatively easy for pupils to read and that words which do not appear on the list will be unfamiliar and more difficult to read.

It is certainly the case that, from analyses of English word usage, a fairly small number of words appear to make up a substantial proportion of words in common use. The graph below is based on data given in Nation and Waring (1997) and derived from the Brown University corpus of present-day English (Francis and Kucera, 1979). The graph plots the percentage of text in English which can be accounted for by numbers of distinct words, counted in thousands. It shows that over 70 per cent of English text is composed of just 1,000 words. Each successive 1,000 words accounts for a smaller and smaller proportion of English text.



One implication of accounting for English word frequency in this way is that minimising the text used in written assessment tasks to the first 1,000 of these words would maximise the readability of these tasks. These 1,000 words are given in an alphabetical list in the Appendix (although it should be noted that this list was derived from a corpus of American English – see below).

One approach along these lines which will be quite familiar to British primary school teachers (of a certain age!) was used as the basis for a very popular UK reading scheme for pupils. McNally and Murray (1968) claimed that while the average adult had a speaking vocabulary of about 20,000 words, an extremely large proportion of the language which people produced, and read, was made up of just 250 words. They then argued that if pupils were systematically taught these 250 words, they would be able to read the vast majority of any text they came across. The Ladybird Keywords reading scheme was devised to focus upon these keywords, which are listed below.

12 most common words in the average vocabulary	a and he I in is it of that the to was
next 20 most common words	all as at be but are for had have him his not on one said so they we with you
next 68 most	about on back been before big by call come can come could did

common words	do down first from get go has her here if into just like little look made make me more much must my no new now off only or our over other out right see she some their them then there this two up want well went who were what when where which will your old
next 150 most common words	after again always am ask another any away bad because best bird black blue boy bring day dog don't eat every fast father fell find five fly four found gave girl give going good got green hand head help home house how jump keep know last left let live long man many may men mother Mr. never next once open own play put ran read red room round run sat saw say school should sing sit soon stop take tell than these thing think three time too tree under us very walk white why wish work woman would yes year bus apple baby bag ball bed book box car cat children cow cup dinner doll door egg end farm fish fun hat hill horse jam letter milk money morning Mrs. name night nothing picture pig place rabbit road sea shop sister street sun table tea today top toy train water

Unfortunately, there are some questions about the validity of the means used to determine lists such as this. Perera (1980) noted that many such lists, especially those which tended to be used in readability formulae, were based on frequency counts done in the United States. But the formulae were still used in Britain, where patterns of vocabulary use were different. A comparison of the revised Spache (1974) list (American) with a British frequency count of pupils' written vocabulary (Edwards and Gibbon, 1973) reveals some discrepancies. Words such as *bonfire*, *doll*, *fairy*, *football* and *mummy* are listed as familiar words in the British list but not the American, whereas words like *cabin*, *candy*, *gift*, *parade* and *neighborhood* are listed as familiar words to American pupils but not British.

It has also been suggested that:

average word frequency is not a good predictor because many words are common at certain age or level, but then become uncommon – such as 'kitten'. But in cases like these, infrequency at higher grade level does not make them difficult words. (Milone, 2008, p. 6)

Another issue is that sometimes words might appear familiar but only with one particular meaning. Homonyms or words used with their less common meanings might confuse the familiarity criterion (Perera, 1980). Nevertheless, the advice given by Ofqual (2010) that "Differentiation should be based on subject content rather than vocabulary" seems sensible given what we know about the enabling effects of the use of familiar words.

## Sentence difficulty

### Principles outlined in the guidance

- Sentence length and sentence complexity can play a part in the readability of assessment questions. But caution is needed when taking these features into consideration:
  - Sentence length is not an absolute guide to sentence difficulty.
  - Sentence complexity can cause misunderstandings in inexperienced readers, and therefore it can be a more useful indicator of sentence difficulty in assessment questions.
- Sentence complexity is influenced by the following features within sentences:
  - The number of propositions (clauses)
  - The number of embeddings
  - The order in which major elements appear
  - The distance between crucial elements.

The common belief regarding sentence difficulty is that the longer its sentences, the harder a text is to read. Hence, the average sentence length in a text has often been used as measure of text difficulty. Most readability formulae have included this as a measure, usually calculating it by dividing the number of words by the number of sentences in a text. Ofqual (2010) recommends that assessment designers:

- use simple sentence structures with a logical conceptual flow (subject, verb, object).
- avoid subordinate clauses.
- present information in short sentences.
- divide even relatively short sentences if they contain a lot of condensed information.

An example of where this advice has plainly been followed occurs in the Key Stage 2 Science Test (2007). The introduction to one question is worded as follows:

*Nadif is growing some plants from seeds. He takes a seed tray and fills it with damp soil. He plants some seeds. Then he puts a transparent lid over the top.*

This could have been worded differently:

*Nadif is growing some plants from seeds. He takes a seed tray, fills it with damp soil, plants some seeds and then puts a transparent lid over the top.*

The second wording would undoubtedly have made the text more difficult to follow and probably have affected the responses to the subsequent questions.

Care needs to be taken, however, in using sentence length as an absolute measure of reading difficulty. Consider the two following text extracts, for example:

Extract 1: *One is, and in the end one is not. That is the basic human problem.*

Extract 2: *The prince jumped on his horse and charged towards the dragon, not noticing the tiny flames which were beginning to appear in its nostrils.*

Extract 1 contains two sentences with an average length of 7.5 words, whereas in Extract 2 there is just one sentence, 23 words in length. Most readability formulae would, on this count, judge Extract 2 to be more difficult to read than Extract 1. But if we are expecting reading to involve some understanding, this is almost certainly not the case for most readers. There is enough evidence that conceptual difficulty is even more important than sentence length in this case.

One of the reasons that longer sentences might be harder to read is that they tend to contain subordinate clauses with complex relationships between them. Nonetheless, there are cases where the same length of sentence brings a different level of difficulty. For example Johnson (1998) suggests that the following sentences will vary widely in difficulty of reading, especially for pupils:

- *The cat sat on your mat.*
- *The cat on the mat.*
- *On the mat the cat sat.*
- *Sat, on the mat: the cat.*
- *The cat on the mat sat.*
- *Sat: the cat on the mat.*
- *Sat the cat on the mat.*

Also Perera (1980) argues that at times longer sentences are easier because they provide more clues as to meaning of the sentence and the relationship between its parts. It is arguable, for example, which of the following is easier to read:

- *The man, who had a wooden leg, sat down quickly.* (10 words)
- *The wooden-legged man sat down quickly.* (6 words)

We can also do a simple experiment to test the notion that the reading difficulty of a text is caused by its sentence length and complexity (or by its word length or familiarity).

Consider the following extract from a famous speech made by Winston Churchill (then British Prime Minister) on 4th June 4th 1940, as Britain was under threat of invasion by the forces of Nazi Germany.

*We shall fight in France, we shall fight on the seas and oceans, we shall fight with growing confidence and growing strength in the air, we shall defend our island, whatever the cost may be. We shall fight on the beaches, we shall fight on the landing grounds, we shall fight in the fields and in the streets, we shall fight in the hills; we shall never surrender, and if, which I do not for a moment believe, this island or a large part of it were subjugated and starving, then our Empire beyond the seas, armed and guarded by the British Fleet, would carry on the struggle.*

This text consists of two sentences, with an average length of 54 words. The following text is a completely reversed version of the first, with the same number of sentences and the same punctuation structure.

*Struggle the on carry would Fleet, British the by guarded and armed seas, the beyond Empire our then starving, and subjugated were it of part large a or island this believe, moment a for not do I which if, and surrender, never shall we hills; the in fight shall we streets, the in and fields the in fight shall we grounds, landing the on fight shall we beaches, the on fight shall we be. May cost the whatever island, our defend shall we air, the in strength growing and confidence growing with fight shall we oceans, and seas the on fight shall we France, in fight shall we.*

Both of these texts were put separately through the readability checker at [www.wordscount.info/](http://www.wordscount.info/). The outcomes are shown in the table below.

<b>Metric</b>	<b>Results for original text</b>	<b>Results for reversed text</b>
Number of words	108	108
Number of sentences	2	2
Number of syllables	138	138
Number of big words [> 3 syllables]	5	5
Smog grade <sup>1</sup>	12.16	12.16
Gunning-Fog Index <sup>2</sup>	23.45	23.45
Automated Readability Index <sup>3</sup>	25.28	25.28
Flesch-Kincaid Grade <sup>4</sup>	20.55	20.55
Spache Index <sup>5</sup>	9.24	9.24

---

<sup>1</sup> **SMOG grade:** Harry McLaughlin (1969) created the SMOG grade, a measure of readability that estimates the years of education needed to understand a piece of writing. The SMOG grade yields a 0.985 correlation with a standard error of 1.5159 grades with the grades of readers who had 100 per cent comprehension of test materials. (McLaughlin, 1969)

<sup>2</sup> The **Gunning-Fog Index:** In 1952, Robert Gunning, an American textbook publisher, developed the Gunning-Fog Index to measure the readability of English writing. The index estimates the years of formal education needed to understand the text on a first reading. A Fog index of 12 requires the reading level of a US high school senior (around 18 years old). The Gunning-Fog Index formula implies that short sentences written in plain English achieve a better score than long sentences written in complicated language.

<sup>3</sup> The **Automated Readability Index (ARI)** is a readability test designed to gauge the understandability of a text. The index produces an approximate representation of the US grade level needed to comprehend the text. (Senter and Smith, 1967)

<sup>4</sup> The **Flesch–Kincaid Grade Formula**, devised by Rudolph Flesh (1948), translates the 0–100 score to a US grade level, making it easier for teachers, parents and others to judge the readability level of various books and texts. The result is a number that corresponds with a grade level. For example, a score of 8.2 would indicate that the text is expected to be understandable by an average student in 8th grade (usually around ages 12–14 in the United States of America). (Flesch, 1948)

<sup>5</sup> The **Spache Readability Formula** is a readability test for writing in English, designed by George Spache. It works best on texts that are for children up to 4th grade. (Spache, 1953)

Dale-Chall Grade <sup>6</sup>	16+	16+
-------------------------------	-----	-----

Incredibly, both texts have exactly the same readability scores, on all the word/sentence measures used. There is a huge gap between their levels of comprehensibility, nevertheless, which gives us some important information about the usefulness of purely using word and sentence measures to determine the likelihood of young readers understanding any text. The likeliest explanation here is that, in judging the language accessibility of a text, the measures we have discussed so far are missing some crucial information.

Nevertheless, there is evidence that sentence complexity can make a difference to the reading comprehension of a text. In a classic study, Reid (1972) took sentences from a range of reading material produced for 7- to 8-year-olds which she judged to be ambiguous and difficult in their syntactic structure. These sentences were then rewritten to make them less ambiguous and the two versions were shown to 7-year-old pupils, who were then asked questions about the sentences. Some examples of Reid's material are as follows:

*Example 1*

Original sentence: *The girl standing beside the lady had a blue dress.*

Modified sentence: *The girl had a blue dress and she was standing beside the lady.*

Question: *Who had a blue dress?*

*Example 2*

Original sentence: *Tom's mother was anything but pleased.*

Modified sentence: *Tom's mother was not pleased at all.*

Question: *Was Tom's mother pleased?*

In Example 1 only 41 per cent of the pupils answered the question correctly after reading the original sentence, and 88 per cent of those who read the modified sentence answered correctly. In Example 2 the percentages of pupils making correct

---

<sup>6</sup> The **Dale-Chall Grade Readability Formula** is a readability test that provides a numeric scale of the comprehension difficulty that readers will have when reading a text. It uses a list of words that groups of 4th-grade American students could reliably understand, considering any word not on that list to be difficult. (Dale and Chall, 1948)

answers were 43 per cent for the original sentence and 80 per cent for the modified sentence.

Reid used her findings to advocate that:

- Greater consideration needed to be given to the linguistic structures used in early reading material.
- Pupils were put at a disadvantage when it came to reading texts such as those she included in her study unless they had had a great deal of prior experience of being read aloud to from texts using similar structures.

This suggestion is probably also true of pupils reading certain sentence structures in assessment questions.

In an experiment to determine the abilities of school-aged readers to understand different levels of sentence complexity, Ecalle and other (2011) used pairs of sentences like those in the table below. In some cases the sentences in a pair have the same meaning; in other cases the second sentence has a complete different meaning. Readers aged from 8 to 15 were asked to read each pair and to indicate whether they each had the same meaning or not.

The snail is neither fast nor noisy.	The snail is slow but not noisy.
The tailor mends the dungarees. He sends the dungarees.	The tailor mends the dungarees and sends them.
The rabbit eats the rat.	The rabbit is eaten by the rat.
The sea lion is less hairy than the chimpanzee.	The chimpanzee is more hairy than the sea lion.
The lozenge which is in the oval is beige.	The lozenge is in the beige oval.
The circle is on the book which is small.	The circle is on the small book.
The cow which the dog looks for is brown.	The brown dog looks for the cow.
The farmers enclose the cattle which the bird of prey observes.	The farmers enclose the cattle and the cattle observe the bird of prey.
The jar in which there is a box is solid.	The box is in the solid jar.

As might have been expected, the readers' performances on this measure improved with age, but even for the 15-year-olds the sentence structures appeared to cause some difficulties. It should be noted here that it was not the content of these sentences which proved difficult, but the complexity with which this content was expressed.

Thompson and Shapiro (2007) have identified four variables that contribute to sentence complexity:

- The number of propositions within a sentence (this equates to the number of verbs, which in turn indicates the number of clauses within the sentence). So, a sentence such as: *The man who was driving the car was in a hurry* is more complex, and hence more difficult to read than: *The man was driving the car. He was in a hurry.*
- The number of embeddings. Thus the following sentences increase in complexity and difficulty:
  - *The man was driving the car.*
  - *The man wearing the sweatshirt was driving the car.*
  - *The man wearing the sweatshirt with the Manchester City logo was driving the car.*
  - *The man wearing the sweatshirt with the Manchester City logo which he had bought from the Internet was driving the car.*
- The order in which major elements appear in the sentence, from simple, active sentences such as subject-verb-object (SVO) to passive sentences (OVS). Thus the following sentences increase in complexity and difficulty:
  - *John (S) kicked (V) the ball (O).*
  - *The ball (O) was kicked (V) by John (S).*
  - *It was the ball (O) that John (S) kicked (V).*
- The distance between crucial elements in the sentence. The basic premise is that when words intervene between elements that are typically closer together in a simply constructed sentence, the reader has to work harder. Notice how in the sentences given above in the second bullet point, the distance between key elements in each sentence (man... was driving) increases from 0 words to 15 words.

## Cohesion and coherence

### Principles outlined in the guidance

- The success of pupils in assessments can be influenced by:
  - the coherence of the texts used to introduce and pose assessment questions
  - the levels of cohesion in these texts.
- Spelling out too carefully the cohesive links between the various parts of a question text might not be a useful approach.
- Cohesive ties, however, must be clear and unambiguous.

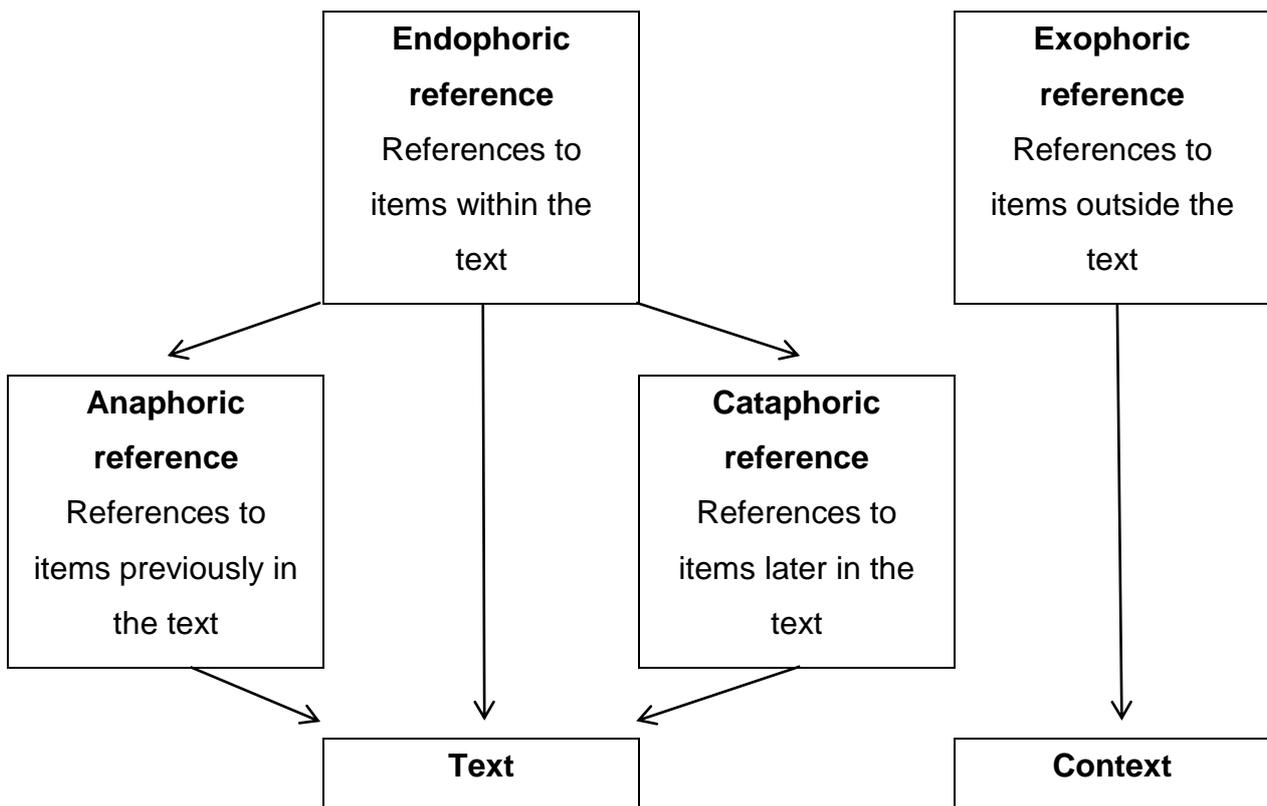
The choice of the words and of the sentences used in an assessment may lead to accessibility issues. But one of the key features of a text is that it is not just a group of words and sentences. Instead, there is a structure in a text which glues the various text components together. In reading a text, the reader needs to construct a coherent, mental representation of the ideas which have been cohesively presented in the text. Louwerse and Graesser (2004) use the term 'coherence' for the way ideas 'hang together' in a text and 'cohesion' for the textual links through which coherent ideas are built up. The effects on readability of the cohesion and coherence of the texts used in assessment questions are often not explicitly considered by test designers.

Yet problems of cohesion can easily cause difficulties for pupils reading assessment questions. The beginning of the Key Stage 2 English (2009) Reading answer booklet, for example, has the following:

*You have now had 15 minutes to read No place like home and The Earthship leaflet. In this booklet, there are different types of question for you to answer in different ways.*

It may well be that some pupils reading this thought initially that there should be questions for them to answer in the *Earthship* leaflet. The reference "this booklet" might well be interpreted to refer to the previously mentioned leaflet, instead of the booklet the pupils are actually reading. The reference is exophoric (to something outside the current text) rather than anaphoric (internal to the text), although anaphoric references are far more common in texts.

Connor (1996) defines cohesion as “the use of explicit linguistic devices to signal relations between sentences and parts of texts”. These cohesive devices are phrases or words that help the reader associate items or statements in a text with others elsewhere in that text, or outside. The following diagram summarises the ways in which such devices work.



Halliday and Hasan (1976) originally identified four general categories of cohesive devices in texts:

- **Reference.** These are the cohesive devices in a text that can only be interpreted with reference either to some other part of the text or to the world experienced by the sender and receiver of the text. Reference items include pronouns (personal reference), demonstratives and the article *the* (demonstrative reference), and items like *such as*, *more*, *as much* (comparative reference). They may be:
  - anaphoric (referring to items previously mentioned in a text: *The ball bounced when the boy kicked it*)
  - cataphoric (referring to items mentioned later: *When he arrived, John noticed that the door was open*)

- exophoric (referring to items outside the text: *Take a look at this*).
- **Substitution and ellipsis.** Whereas reference indicates a meaning relationship between two items, substitution is more grammatical in nature. A word, phrase or clause is substituted in a following sentence for one with a similar grammatical function. Look at the following short text:

*Car tyres eventually wear out, of course. New ones have to be fitted.*

*Ones* here is used as a substitute for *tyres*.

Another type of cohesive tie which operates in very similar ways to substitution is ellipsis. Here a word or phrase is missed out from a text after it has been mentioned once. Look at the following:

*The postman walked along the road and delivered letters to every house.*

*The postman* is implied but not stated before the word *delivered*. This phenomenon is known as ellipsis and, as can be seen in this example, appreciating how it works is crucial to understanding the text. The reader has to be able to supply, almost sub-consciously, the missing word to make sense of the sentence.

- **Lexical.** Lexical cohesion occurs when two words in a text are semantically related in some way – in other words, they are related in terms of their meaning. Words might be repeated or, more commonly, a synonym used. Or pairs of words might naturally occur together – *fish and chips*: such pairings are known as collocations.
- **Conjunction.** Conjunction differs from reference, substitution and ellipsis in that it does not set off a search backward or forward for its referent. However, it is a linguistic cohesive device in that it signals a relationship between segments of the text, which might be additive (*and*), causal (*because*) or temporal (*then*).

A text may be cohesive without necessarily being coherent. Cohesion relies on lexical and grammatical relationships, but coherence is based on semantic relationships. If cohesion does not automatically guarantee coherence then neither is the reverse relationship true. Look at the following example:

*Jill: The phone's ringing.*

*Jack: I'm tired.*

In this case, there are no explicit cohesive markers to bind these two sentences together. It seems that Jack has totally disregarded, or failed to interpret, the

meaning of Jill's utterance. As readers, we naturally assume that this sequence of sentences does constitute a text and we interpret the second sentence in the light of the first sentence. Perhaps Jack's reply indicates that he feels he always has to answer the phone and wants Jill to do it on this occasion. Or perhaps Jack knows who is on the other end of the phone line and does not want to talk to the person this late.

These seemingly unconnected sentences have been made to form a coherent text, but only by the reader supplying 'real-world knowledge'. Making sense of any text involves interpretation and depends to a great extent on what the reader brings to the text. The reader has to rebuild the world of the text, see into the mind of the writer, using her/his experience of that world. The reader has to activate her/his background knowledge, make inferences and constantly re-interpret as new information is provided.

Studies of cohesion in reading show that it can make a substantial contribution to readability. One study (Chapman, 1987) demonstrated that readers between the ages of 8 and 15 showed growth in their ability to perceive cohesion in text and to use it to support their comprehension. This suggests that readers develop an awareness of cohesion over time and make increasing use of it to get meaning from print. However, having insufficient experience and knowledge of the ways in which texts are cohesive and coherent can be a major hindrance to their comprehension. Other studies, such as that by Fulcher (1989), have suggested that readers' failure to comprehend a text can result from their inability to follow the flow of cohesive ties within the text. A more complex picture, though, is provided by the research of Ozuru and other (2009). They compared the reading of science texts which were deliberately written to have either high or low cohesion between sentences. An example of each of these types of text is given below:

*Example 1. Heat distribution in animals (low cohesion)*

The circulatory system is responsible for the distribution of heat throughout the body. This is true for both warm-blooded and cold-blooded animals. The term 'warm blooded' is applied to birds and mammals in recognition that they can, and usually do, keep their body temperature higher than that of their surroundings. But this is not always the case; some of them allow their temperature to drop close to the ambient temperature, when they hibernate, for example. And some of them, mammals in the tropical savannah, for example, have to keep their body temperature below the scorching temperatures of the surroundings. However, there are two features that set birds and mammals apart from most of the rest of the animal kingdom:

- They maintain their body temperature within narrow limits no matter what the ambient temperature. For this reason, they are often described as being homeothermic.
- They are endothermic; the heat with which they maintain their body temperature is generated within the body. Some coldblooded animals, for example, lizards basking in the sun, develop body temperatures as high as that of birds, but they are ectothermic; they secure the heat for doing so externally.

*Example 2. Heat distribution in animals (high cohesion)*

The circulatory system *distributes heat through the blood vessels of an animal's body. This system is responsible for the transport of heat for both warm-blooded animals and cold-blooded animals. Warm-blooded animals include birds and mammals, whereas cold-blooded animals include reptiles, amphibians, and fish. The term 'warm blooded' is applied to birds and mammals because they can, and usually do, keep their body temperature higher than that of their surroundings. But this is not always the case because some warm-blooded animals allow their body temperature to drop close to the temperature of the air around them, for example, when they hibernate through the winter. Mammals who live in the heat of the tropical savannah *are another example of warm-blooded animals that do not always keep their body temperature higher than the surrounding temperature. These animals often have to keep their body temperature below the scorching temperatures of their surroundings.**

Nonetheless, there are two features that set warm-blooded animals apart from most of the rest of the animal kingdom:

1. Warm-blooded animals are homeothermic. That is, unlike other animals, birds and mammals maintain their body temperature within narrow limits no matter what the surrounding (or ambient) temperature.
2. Warm-blooded animals are endothermic; that is, they maintain their body temperature with heat generated within their own body.

Endothermic animals contrast with cold-blooded animals whose body temperature is maintained by heat from external sources. As such, even though some cold-blooded animals, such as lizards who bask in the sun, develop body temperatures as high as that of birds, these creatures secure their body heat externally. These kinds of animals are called ectothermic.

*Note: The sections underlined were added to increase cohesion. The italics sections indicate where changes in sentence structure were made to increase cohesion.*

This study found that the effect of text cohesion depended both on the reading skill and the level of prior knowledge of the reader. Higher text cohesion seemed to benefit readers with poorer levels of prior knowledge. However, readers with lower levels of reading skill but higher levels of prior knowledge of the topic of a text tended to process the text more shallowly and actually perform less well on a subsequent assessment of their understanding. This finding replicates that of O'Reilly and McNamara (2007) and suggests that readers' difficulty in learning new concepts can be alleviated to some extent by making text more cohesive, which makes readers less dependent on pre-existing knowledge.

Yet, it seems that readers are not able to take advantage of increased cohesion unless they have sufficient reading skill. It may well be important for teachers not only to work on improving pupils' understanding of content, but also on their abilities to read to learn from texts. Teachers may also promote pupils' skills to read assessment questions more effectively.

A further implication relates to the need to improve the texts pupils are asked to read for both learning and assessment purposes (Beck and other, 1991; Graesser and other, 2003). Such texts need to be evaluated for their levels of cohesion. But it should not be taken for granted that increasing the levels of cohesion – for example by spelling out all the cohesive links within a text – will benefit readers in the same way. Readers who are knowledgeable about the topic of a text they are reading for learning or for assessment may actually be helped to show this knowledge if the text is less cohesive – that is, makes greater use of reference links, especially anaphora.

## Content structure and complexity

### Principles outlined in the guidance

- The success of pupils in assessments can be influenced by the content structure of the texts used in the assessment.
- Features for assessment developers to be aware of in terms of content structure include:
  - The need for unambiguous information
  - The propositional density of the text used, which must not be so high as to overwhelm the pupils' reading and comprehension capacities
  - Possible lexical incoherence issues and the need to ensure that the various parts of a text (phrases, sentences, paragraphs) clearly link together.

Well-written text requires, in addition to coherence and cohesion, a structure that readers can easily use to find the information they need and then to understand it correctly. Text can become confusing when information is inappropriately presented. Most sentences, when taken out of context, become ambiguous in meaning. When we read text, we build a collection of the concepts described therein, deducing these concepts from the words and phrases used within the text. We build certain interpretations out of these blocks of words which are not randomly organised, but obey quite strict rules of association. For example, words and their meanings impose restrictions on possible synonyms – *strong tea* may be acceptable, but *powerful tea*

probably is not (although whisky can be either strong or powerful!). A reader unfamiliar with such constructions might not understand the precise meanings in the text.

When linguistic expressions combine into units for processing, many of the individual linguistic elements are ignored and the whole chunk is treated as one semantic unit. When a significant amount of information is conveyed in a relatively small amount of text, the reader can easily become confused. This problem is known as ‘propositional density’ (Kintsch, 1974). The greater the number of ideas expressed in a text, the more work is required of the reader to interpret the text correctly (Newbold and Gillam, 2010).

Kintsch and Keenan (1973) presented readers with sentences of constant length but varying propositional density. They found that as the number of propositions in a text increased, so did both the time taken by readers to read the text and the number of propositions they were able to recall from the text. This suggests that the unit of meaning that readers deal with in reading is the proposition. It was also the case that readers remembered high-level propositions more than low-level ones, which suggests a hierarchical approach to understanding a text. However, there is now quite robust evidence that high propositional density in a text adversely affects readers’ understanding of that text (for example Barshi and Healy, 2002; Sonnleitner, 2008). The implication for assessment designers links with the earlier recommendations about sentence complexity in assessment questions. The more complex and propositionally dense the text of a question, the harder that question is to answer, no matter the pupil’s actual content knowledge.

The second problem with text structure is called ‘lexical incoherence’. This occurs when writers present new information to the reader without making clear its relationship to previous information. If a large number of new, seemingly unrelated ideas are introduced then a reader can find it very difficult to make sense of a text (Newbold and Gillam, 2010).

Sometimes writers use apparent lexical incoherence to create interest in a text. As an example, look at the following extract from a letter by the mother of Norman, the main protagonist in *Dear Norman* (the reading text used in the Key Stage 2 English Reading paper, 2009):

*By the way, did you see those sweet busy bees below your tree house? I suppose they could be hornets or wasps, but don’t worry, they look more like jolly bumble bees to me.*

On the face of it, these bees have nothing to do with the story unfolding in this book, of Norman leaving home to live in his tree house. The reader has to interpret the mother’s motives for introducing this information before the point of the bees can be

grasped. For several readers of this text, it is quite likely that this interpretation was never made.

## **Legibility and print issues**

### **Principles outlined in the guidance**

- The success of pupils in assessments can be influenced by the legibility of the texts used in the assessment.
- Assessment developers need to:
  - Carefully consider the layout of written assessments, because poor presentation can hinder effective communication. Judicious use of white space seems to be the key here, to avoid the layout looking cluttered.
  - Use an appropriate font of sufficient size (at least 12 point but 14 point is better).
  - Use underlining, bold, italics, boxes, indentation and shading consistently.
  - Use headings, subheadings, bullet points and numbers to ensure questions are well structured, clear and easily managed.
  - Use simple rubric in plain English so that the question or task is clear.
  - Use diagrams, pictures or photographs to convey key information graphically as well as textually, but only when there is a clear purpose or benefit to the majority of pupils.
  - Avoid questions where the correct answer requires pupils to distinguish between different colours. This could disadvantage colour-blind pupils.
  - If answer booklets are being used, provide enough space for pupils' responses.
  - Clearly show the mark allocation for each question or question part.
  - Ensure cover pages are clearly laid out and include only essential information about the assessment.

Tinker (1963) defines legibility as “concerned with perceiving letters and words, and with the reading of continuous textual material” (cited in Lund, 1999, p. 17). Legibility studies have usually researched factors such as size of characters, thickness of

strokes, white space between strokes, dissimilarity of characters, leading, line length, quality of paper, colour of paper and colour of ink (Waller, 1991, p. 342). Research has shown that legibility issues such as the size of font and typeface can affect reading and reading speed (Hughes and Wilkins, 2000; Wilkins et. al, 2009).

Ofqual (2010) includes several examples of legibility issues which are likely to make a difference to text readability, such as font style, size and weighting. The advice given is that assessment developers should:

- carefully consider the layout of written assessments, because poor presentation can hinder effective communication. Judicious use of white space seems to be the key here, to avoid the layout looking cluttered.
- use an appropriate font of sufficient size.
- use underlining, bold, italics, boxes, indentation and shading consistently.
- use headings, subheadings, bullet points and numbers to ensure questions are well structured, clear and easily managed.
- use simple rubric in plain English so that the question or task is clear.
- include diagrams, pictures or photographs in questions only when there is a clear purpose or benefit to all pupils (although of course there will be some visually impaired pupils for whom the inclusion of these graphic elements will never be of benefit).
- avoid questions where the correct answer requires pupils to distinguish between different colours. This could disadvantage colour-blind pupils.
- if answer booklets are being used, provide enough space for pupils' responses.
- clearly show the mark allocation for each question or question part.
- ensure cover pages are clearly laid out and include only essential information about the assessment.

Research tends to support this advice. Eyles and other (2003) found, for example, that a sans-serif font was generally a preferable typeface to ease readability (see also Wilkins and other, 2009), although it has sometimes been argued that serif fonts ease reading because the serifs draw the eye along the line.

Text legibility is also influenced by the size of the font (see, for example, Feely and other, 2005; Pillai and other, forthcoming; and Wilkins and other, 2009). Studies have shown that by increasing the font size, the percentage of fluent reading is also increased (Feely and other, 2005) and small font sizes (below 12 point) are thought

to make reading increasingly difficult, and are more stressful to the visual system (Wilkins and other, 2009).

The key features in the legibility of various text display options are summarised by White (2004). These are listed in the following table:

<b>Text display characteristic</b>	<b>Legibility features</b>
Font size	<p>Optimal font size is between 10 pt and 15 pt. Smaller type becomes less legible. Compare the following:</p> <p>How easy is this to read? (10 pt)</p> <p>How easy is this to read? (12 pt)</p> <p><b>How easy is this to read? (18 pt)</b></p>
Font weight	<p>Medium-weight fonts are the easiest to read. Bold fonts attract attention in comparison. Compare the following:</p> <p>How easy is this to read? (Arial Narrow)</p> <p>How easy is this to read? (Arial)</p> <p><b>How easy is this to read? (Arial Black)</b></p>
Font style	<p>Roman is easier to read than italic. Italic can be used to emphasise short blocks of text for greater attention. There has been debate about the merits of serif and sans-serif fonts. Compare the following:</p> <p><i>How easy is this to read? (Italic)</i></p> <p>How easy is this to read? (Sans-serif)</p> <p>How easy is this to read? (Serif)</p>
Line length	<p>Roughly 50 to 65 characters is an ideal measure. Anything significantly smaller or larger loses legibility.</p>
Letter spacing	<p>Most fonts do not need extra letter spacing (leading) because it should be part of the font design. It is best to keep with the original design.</p>

Word spacing	<p>The space between words should be large enough to indicate clearly that they are different words, but not so large as to lose their connection to each other. Compare the following:</p> <p>How easy is this to read? (single word spacing)</p> <p>How easy is this to read? (triple word spacing)</p>
Line spacing	<p>The vertical space between lines of type should increase as the length of the lines increases. The lack of white space between lines makes it difficult for the eye to track from one line to the next. Usually a line-height between 1.15 and 1.5 works well.</p>
Justified or flushed left or right	<p>Type that is set flush left with a soft right edge is the easiest to read. Full justification risks creating uneven spaces between words. Compare the following:</p> <p>These lines are flushed left and have an unjustified right edge. Their advantage is that they maintain the same distance between words throughout.</p> <p>These lines are fully justified – that is, straight on both left and right sides. It will be apparent that the between-word spaces are not equal here, which can make reading harder.</p>
Lowercase, all caps, small caps	<p>Lowercase letters have more contrast in their strokes, making them easier to read. ALL CAPS should be reserved for display type. SMALL CAPS can be used for emphasis, but sparingly.</p>
Contrast between type and background	<p>Black text on a white background is the most legible. Everything else reduces legibility.</p>

## **Text organisation**

### **Principles outlined in the guidance**

- The success of pupils in assessments can be influenced by the organisation of the texts used in the assessment.
- Consideration needs to be given to the use of:
  - format variables, including paragraphing, headings and subheadings, and typographical effects such as bold type, underlining and italics.
  - advance organisers; that is, summarising pieces at the beginning of longer texts to alert the reader to what is coming and to help enhance his/her conceptual organisation.

Harrison (1984) has suggested that there three aspects of text organisation are important in educational contexts:

- format variables, which include paragraphing, headings and subheadings, and typographical effects such as bold type, underlining, and italics (there are findings (Waller, 1991) which suggest that format variables such bold type, underlining and italics can be confusing or even distracting rather than helpful to readers)
- advance organisers: summarising pieces that are included at the beginning of longer texts to enhance the reader's conceptual organisation
- the use of printed questions as a part of books that are designed to promote learning and understanding.

## Readability: looking at the characteristics of readers

### Principles outlined in the guidance

- The readability of assessment questions is determined as much by characteristics of the pupils as it is by features of the texts themselves.
- Differential Item Functioning (DIF) analyses have been carried out extensively to try to determine the impact which particular items may introduce into assessments for particular groups of pupils.<sup>1</sup>
- The aim is to produce assessment questions which are ‘fair’; that is, which have no inbuilt bias for or against particular groups of pupils.

It is unlikely that two pupils are going to perform exactly the same when faced with a test, especially one which involves the extensive interpretation of written language. Pupils are not clones of one another and each has his/her own individual characteristics which affect, however slightly, his/her responses to assessment questions. As mentioned previously, an important omission in most research into readability is the effect of various reader characteristics. We now recognise that readability and language accessibility are both products of the features in a text **and** the characteristics of a reader. Therefore it is important for writers of assessment questions to take into account these reader characteristics if they are to work towards what Cole and Zieky (2001) have termed “the new faces of fairness”.

Understandings of the ways in which reader characteristics can affect the readability of assessment questions have been developed over a number of years through the use of Differential Item Functioning (DIF) analysis (Gierl, 2005; Zumbo, 2007). This well-established statistical procedure has been used to identify individual questions in assessments that may be biased against particular groups of pupils.

In a DIF analysis the performance on each question of all the members of one group of pupils is compared with the performance of the members of another group. For example in a gender-based DIF analysis the results for girls and for boys might be compared for each question in a test. This can help assessment developers to identify particular items on which members of one of these groups perform in a way that does not match their overall performance on the assessment. So, for instance, a gender-based DIF analysis might reveal that girls who perform well on the

---

<sup>7</sup> For a comprehensive approach to Differential Item Functioning (DIF) refer to *Did it work? Evaluating access to National Curriculum assessments. Guidance and Research Background*, Ofqual 2012.

assessment overall tend to perform less well on a specific question or part of a question. In this case the question needs to be reviewed to check that it does not have some hidden barriers to accessibility for girls – question topic, for example.

Gierl (2005) has suggested four key aspects to test fairness:

- freedom from bias – that is, not producing outcomes that unfairly favour or disadvantage members of particular groups
- ensuring that pupils receive equal treatment in the testing process
- promoting equity in the outcomes of assessment – that is, a particular outcome should have similar consequences for all pupils
- giving pupils the opportunity to learn the content covered in the assessment.

Bias occurs when assessments produce different scores or promote different score interpretations for members of different groups (for example groups with differences in racial, ethnic, language, cultural, gender, disability or socio-economic statuses).

DIF analysis now has a substantial history and large numbers of research studies have been carried out using its principles and methods. But we still lack a full understanding of just why DIF occurs in educational assessment (Gierl and other, 2003). To develop such an understanding requires an appreciation of the kinds of pupil characteristics that have been shown to affect the readability of the texts used in assessments. These characteristics will be examined in the following sections.

### **Principles outlined in the guidance**

- Characteristics of pupils which have been shown to have an influence upon their reading and understanding of assessment questions, and hence on their success in assessments, include the following:
  - Physical capabilities – special educational needs such as autism, dyslexia and ADHD, and physical issues such as visual or hearing impairments can all influence performance in assessments.
  - Reading abilities – difficulties in reading are likely to significantly affect a pupil's performance in a test, no matter what level of content knowledge he/she may have of that subject.
  - Engagement/motivation – one of the most powerful explanations for pupils' performance in assessments is their interest in the content of these assessments and/or their emotional reaction to this content.
  - Prior knowledge – this influences not just pupils' demonstration of content knowledge of a test, but also their approach to reading the test questions. Language and cultural background are particularly strong influences on assessment performance.
  - Gender – gender differences in assessment responses have been commonly found in assessment research. But it should be remembered that there may be greater variation in responses within a gender than between genders.

## **Physical capabilities**

Pupils start learning language from birth and it becomes more complex as they grow up. At preschool most pupils learn the letters of the alphabet, as well as the letter sounds. Clearly, having a disability or impairment can influence the progression of a pupil's reading ability.

Some examples of impairment that can lower a pupil's reading level include autism, dyslexia and ADD (Attention Deficit Disorder) and ADHD (Attention Deficit Hyperactivity Disorder). Pupils with ADD and ADHD<sup>8</sup> have difficulties in concentrating on a task for any lengthy period of time. Autistic pupils need special teaching techniques because they are often unable to interact with others. Dyslexia affects pupils' reading ability in that it might be difficult for them to translate images to language and this may cause difficulty in spelling and reading (Just and Carpenter, 1987).

There is also a range of physical capability issues which may affect readers: that is, readers who are dyslexic, who have specific learning difficulties or who have hearing or visual impairments. Such issues are likely to have an even greater impact upon the accessibility of assessment texts for younger readers because a child's learning difficulty may not have been identified as yet.

## **Reading abilities**

Reading abilities enable the reader to:

- read meaningful language
- read any written form with independence, comprehension and fluency
- mentally interact with the message from the written form. (Just and Carpenter, 1987; Downing and Leong, 1982)

Hence, the reader needs to master:

- reading skills such as word attack skills, which enable the reader to convert graphic symbols into intelligible language
- comprehension skills that help the reader to comprehend the meaning of print
- various other reading skills.

---

<sup>8</sup> National Institute of Neurological Disorders and Stroke – NINDS Attention Deficit – Hyperactivity Disorder Information page - [www.ninds.nih.gov/disorders/adhd/adhd.htm](http://www.ninds.nih.gov/disorders/adhd/adhd.htm) .

It is obvious that if a pupil is disadvantaged by lack of reading ability then he/she will be much less likely to process any form of text which involves reading, whatever the level of content knowledge he/she may have.

Abedi and other (2008) have confirmed that pupils with disabilities tend to perform in assessments at lower levels than those without disabilities. While their lower performance can be partly explained by the specific disability, there may be other factors that potentially interfere with this performance. It would be useful to identify such factors and attempt to reduce their interference, so that we might get a more accurate picture of the capabilities of pupils with disabilities.

Research by Morgan and other (2008) suggests that what has become known as the “Matthew effect” (Stanovich, 1986) – that is, a pattern of increasing advantage or disadvantage in reading skill development following an initial advantage or disadvantage (“the rich get richer, the poor get poorer”) – is very evident in test-taking situations. Poorer readers are less able to access effectively the written language through which they are tested, and thus demonstrate lower abilities. This causes expectations about their achievements, and perhaps also the level of material upon which they are tested, to be depressed even further.

## **Engagement/motivation**

Engagement or motivation in reading refers to the intrinsic drive to read for the knowledge and the enjoyment that it provides (Guthrie and Cox, 2001). Engagement in reading is important because it drives the reader to use his/her best strategies for understanding and interpreting the text (Guthrie and other, 1997).

There are many examples in the literature and in common experience of readers who can read beyond their normal levels when they are engaged and motivated by particular texts. It was also made clear by the 2001 PIRLS international comparative study<sup>9</sup> that while pupils in all countries had generally positive attitudes toward reading, those with the most positive attitudes had the highest average achievement. It is well founded from a range of research studies that motivational factors have an influence upon their understanding of text. Motivational factors include:

- pupils’ positive and high self-perceptions about their own reading abilities
- the value pupils place on reading
- the enjoyment they derive from reading

---

<sup>9</sup> [www.iea.nl/pirls2001.html#c316](http://www.iea.nl/pirls2001.html#c316)

In addition, motivational factors become more and more important as predictors of the ability to read for understanding as readers get older and develop their skills (Saarnio and other, 1990).

Research on test motivation suggests that this could be a crucial factor in obtaining high-quality and accurate information from assessments in a range of subjects. One study found that test-taking motivation was positively related to subsequent performance on a cognitive ability test, even after the effects of race and performance on the first test were controlled (Chan and other, 1997). Another study found that the validity of a particular test was much higher for a group with more positive motivation towards test-taking than for a group with less positive motivation (Schmit and Ryan, 1992).

As discussed above, the underlying substantive reasons for the differential functioning of some items in assessments are still speculative only (Roussos and Stout, 1996). But one of the most common and widely discussed explanations is pupils' interest in the content of assessments and/or their emotional reaction to this content. Stricker and Emmerich (1999) suggested that both of these explanations could account for the different levels of responses to the assessment questions in their study. Engagement and motivation are equally important in test-taking contexts as in reading and comprehension.

## **Prior knowledge**

Prior knowledge is an integral part of the comprehending process (Johnston, 1984). Hence, prior knowledge influences what is understood from text. This means that two individuals with different prior knowledge but equal levels of reading comprehension skills would still exhibit different levels of comprehension of the same text.

Not surprisingly, pupils who know more about a topic understand and remember content better than those who have a limited background in the domain (Chi, 1985). This factor also comes into play during test-taking. Ozuru and other (2009) found, for example, that while understanding of a science text, as measured by performance on a set of assessment questions, was positively affected by both the reading skill and the prior knowledge of the readers, prior knowledge was a much more significant predictor of test success. This finding supports that of Bugel and Buunk (1996) who claim from their study that the differences which are often found between male and female success in assessments involving reading comprehension (see later discussion) can largely be accounted for by differences in the prior knowledge that each gender tends to bring to the assessment situation.

One aspect of prior knowledge which has been extensively investigated is knowledge of the language of the assessment. Research conducted by Abedi and his colleagues has demonstrated that there is indeed a substantial link between pupils' English

language proficiency and their performance in assessments (in English) of mathematics, science and social studies (for example, Abedi and other, 2003; Bailey, 2000). Furthermore, several studies have found that assessments and individual test questions that are more linguistically complex produce larger performance gaps between pupils of English as an additional language (EAL) and native English speakers (for example Abedi and other, 2003; Abedi and other, 2000). These findings suggest that assessments in all subjects assess language skills as well as content knowledge and skills.

Butler and Stevens (1997) have suggested a number of possible responses to the problems caused by the language of assessments for EAL pupils. These range from modifications of the assessment for these particular pupils (for example carrying out assessments in pupils' native languages, or modification of the language used in test directions), to modifications in assessment procedures for this group (including, for example, extra assessment time or oral directions given in the native language). But a meta-analysis (Kieffer and other, 2009) of studies of the effects of several of these 'accommodations' has proved disappointing, finding little evidence that the assessment performance of EAL pupils is much improved by them (with the possible exception of providing pupils with English dictionaries as they undertook the tests).

What seems more important is to provide EAL pupils with "targeted, explicit, and intensive instruction in the complex and specialised language that lies at the heart of each content area" (Kieffer and other, 2009, p. 1190). However, in a study in the USA, Robinson (2010) did find that Spanish-speaking English language pupils performed significantly better on mathematics assessments when they were tested in Spanish (instead of English). Test translation would be a radical solution to the problem of test access for EAL pupils. But there is, as yet, little evidence that it would be of benefit, and no studies have been reported of the effects of such translation on test outcomes in the UK.

Prior knowledge also includes the social and cultural backgrounds of pupils. The schema theory of reading comprehension proposes that the organisation of prior knowledge in a pupil's mind provides a framework (ideational scaffolding) which enables him/her to understand the setting, mood, characters and chain of events in a text. Readers acquire meaning from a text by analysing the words and sentences against the backdrop of their own personal knowledge of the world. Such personal knowledge is conditioned by a variety of factors: age, gender, ethnicity, nationality, experiences and so on, which make up a person's culture. Readers who share the cultural background of the writer of a text 'come equipped' with the appropriate schemas for making sense of this text. Those who are reading a text based on an unfamiliar culture must pay attention to the details related, but must also try to reach some understanding of the framework underlying these details. The absence of an

appropriate schema might be expected to lead to misunderstandings, which could be very significant in a test situation.

## Gender

Dorans and Kulick (1983) present an example from an American test of general cognitive skills. Pupils were presented with the word pair: *decoy: duck*. They were then asked to choose one of five other word pairs (*net: butterfly; web: spider; lure: fish; lasso: rope; and detour: shortcut*) in which the words had a similar relationship to one another. Using the DIF statistical method, Dorans and Kulick found that pairing related words was more difficult for females than males when overall ability was controlled. They attributed this to gender-related differences in background knowledge, “as it required some knowledge of hunting and fishing, two traditionally male-oriented recreational activities” (p. 20).

Gender differences in test responses have been commonly found in assessment research. Hamilton (1998), for example, conducted research into gender differences in science achievement tests. She found that male pupils were advantaged by the content of the tests, particularly where they were required to bring to bear their existing, out-of-school knowledge. But it was the format of the assessments which gave them the greatest advantage, with the use of diagrams as ways of tuning in pupils to the demands of a question being particularly salient. Gierl and other (2003) produced similar findings in their research into differential gender achievement in mathematics tests, where males did much better than females on questions requiring spatial processing, rather than simple memorisation.

However, although there is documented evidence of gendered differences in reading achievement, as well as attitude, choice and response for some boys (for example Millard, 1997), considerable observable evidence also suggests that this is not the case for all boys. Maccoby’s (1990, p. 513) synthesis of decades of research on gender differences led her to claim that even when consistent differences between males and females were found, the amount of variance accounted for by gender was small, relative to the amount of variation within each gender. It has been repeatedly pointed out that boys are more different than alike, and that statistics lose sight of individual differences. As with all research evidence regarding the effects of group differences in education, we need to be very wary of assuming that all individuals fit the characteristics of the groups to which they belong.

## Accommodations, modifications and universal design

### Principles outlined in the guidance

- Accommodations and assessment modifications have been widely used to cater for pupils with a variety of characteristics. It is not altogether certain how effective these have been.
- An alternative approach, currently used in the development of National Curriculum assessment materials, is to focus on universally designed assessments. Such assessments should be/have:
  - inclusive assessment population
  - precisely defined concepts
  - accessible, non-biased assessment questions
  - amenable to accommodations
  - simple, clear, and intuitive instructions and procedures
  - maximum readability and comprehensibility
  - maximum legibility.

The list of catalogued readers' characteristics is constantly increasing. This ever-widening set of reader characteristics has a significant effect on pupils' demonstration of their capabilities in assessments in a range of subjects. The traditional response of test development agencies, both in the UK and the US, has been to explore various assessment accommodations. Suggested accommodations have included modifications of assessments for particular pupils and modifications in assessment procedures for particular groups.

The table below gives some examples of accommodations in both these categories.

<b>Modifications of assessments for particular pupils</b>	<b>Modifications in assessment procedures for particular groups</b>
Assessment in native language rather than in English	Extra assessment time

Changing (simplifying) the vocabulary used	Breaks during testing
Modifications of linguistic complexity	Administration over several sessions
Addition of visual supports	Oral directions given in the native language
Use of glossaries in native language	Small-group administration
Use of glossaries in English	Separate room administration
Linguistic modification of test directions	Use of dictionaries
Provision of additional example items/tasks)	Reading aloud of questions
	Answers written directly into test booklets
	Directions read aloud or explained

A great deal of research has been carried out to explore the effects of such accommodations. However, such research has proved difficult to conduct and has rarely provided conclusive evidence about the effects of accommodations on assessment validity (for example Stone and other, 2010).

One example is the study of Abedi and other, 2010). Given what we already understand about the difficulties posed to pupils by the complexity of the sentences in the assessment questions, one potential change would be to simplify these sentences by some kind of text segmentation. While a review of literature by Rasinski (1990) did suggest that organising text into smaller units could facilitate memory recall and improve comprehension for certain readers, Abedi and other (2010) found that this made no difference at all to the assessment scores of the pupils with disabilities that they studied.

Thompson and other (2004) have argued for a more global approach to the issue and a move towards universal design in assessments – that is, the design and development of assessments that:

- allow the participation of the widest range of pupils
- produce valid outcomes reflecting the true capabilities of everyone who takes them.

Thompson and other (2004) outline seven key elements which underpin the concept of universally designed assessments.

- **Inclusive assessment population.** Assessments designed for national use must try to include every pupil. They need to be responsive to growing demands – increased diversity, increased inclusion of all types of pupils in the general curriculum, and increased emphasis and commitment to accountability for all pupils.
- **Precisely defined concepts.** The specific constructs tested must be clearly defined so that all irrelevant barriers can be removed. An important function of well-designed assessments is that they actually measure what they are intended to measure. Test developers need to examine carefully what is to be tested and design items that offer the greatest opportunity for success within those constructs.
- **Accessible, non-biased assessment questions.** Accessibility should be built into assessment questions from the beginning, and bias review procedures need to ensure quality in all items. Most importantly, items must be developed by individuals who understand the varied characteristics of the pupils they are aimed at, and the characteristics of items that might create difficulties for any group of pupils.
- **Amenable to accommodations.** The assessment design should facilitate the use of essential accommodations. Although items on universally designed assessments will be accessible for most pupils, there will always be some who continue to need accommodations. For example, the use of Braille as an accommodation will be facilitated if the following features are avoided in the design of the assessment:
  - Use of irrelevant graphics or pictures
  - Use of vertical or diagonal text
  - Items that include distracting or purely decorative pictures, which draw attention away from the item content.

These features are also relevant for pupils with visual disabilities who do not use Braille, and possibly also for the many for whom visual features may create distractions.

- **Simple, clear and intuitive instructions and procedures.** All instructions and procedures should be simple, clear and presented in understandable language. Assessment instructions should be easy to understand, regardless of a pupil's experience, knowledge, language skills or current concentration level.

- **Maximum readability and comprehensibility.** Plain language guidelines should be used to produce readable and comprehensible text. Plain language has been defined as language that is straightforward and concise. Listed below are several strategies that have been identified for editing text to produce plain language:
  - Reduce excessive length by reducing wordiness and removing irrelevant material.
  - Avoid unusual or low frequency words and replace these with common words – for example, replace *utilise* with *use*.
  - Avoid ambiguous words – for example, *crane* should be avoided because it could be a bird or a piece of heavy machinery.
  - Avoid words with particularly unusual or irregular spelling patterns – for example *trough* and *feign*.
  - Avoid proper names and replace with simple common names such as first names.
  - Avoid inconsistent naming and graphic conventions by avoiding multiple names for the same concept and inconsistencies in the use of font.
  - Avoid unclear signals about where pupils' attention should be directed by using well-designed headings and other graphic features (bold, italic fonts) to convey information about the relative importance of information and the order in which it should be considered.
  - Mark all questions clearly by the use of an obvious graphic signal (for example bullet, letter, number) to indicate separate questions.
- **Maximum legibility.** Legibility is the physical appearance of text, the way that the shapes of letters and numbers enable people to read text easily. Bias results when assessments contain physical features that interfere with a pupil's focus on, or understanding of, the constructs that the questions are intended to assess.

## **Conclusion**

As we argued earlier in this review, the concept of readability has developed over the past 20 or so years, in line with theories about the nature of the reading process. Traditionally, studies of readability have focused largely on features in the text itself. In this paper we have reviewed the major conclusions which can be drawn from this line of research: that the readability of a text is influenced by

- linguistic issues such as word and sentence difficulty
- cohesion and coherence
- conceptual difficulty
- legibility and print issues
- text organisation.

More recently, the role of the reader and the readability of texts has become prominent in studies of factors which may affect text comprehension. That is why this guidance has reviewed the influence of such factors as readers' physical capabilities, reading abilities, engagement/motivation, prior knowledge and gender.

There is a strong evidence to support a modern concept of readability that takes into account both the role of the reader and readability of texts side by side with the features of a text (linguistic issues, cohesion and coherence, conceptual difficulty, legibility and text organisation). A major interest for researchers in test design, and for practitioners involved with pupils' testing, is the way in which the features of a text and a reader's characteristics might interact with each other. Test developers and designers need to understand the principles explored in this document if they are to produce 'fair access by design' for all pupils.

## References

- Abedi, J., Kao, J., Leon, S., Mastergeorge, A., Sullivan, L., Herman, J. and Pope, R. (2010) Accessibility of Segmented Reading Comprehension Passages for Students with Disabilities. *Applied Measurement in Education*, 23: 2, 168–186.
- Abedi, J., Leon, S. and Kao, J. C. (2008) *Examining Differential Item Functioning in Reading Assessments for Students with Disabilities (CRESST Report 744)*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Abedi, J., Leon, S. and Mirocha, J. (2003) *Impact of Student Language Background on Content-based Performance: Analyses of Extant Data (CSE Technical Report 603)*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Abedi, J., Lord, C., Hofstetter, C. and Baker, E. (2000) Impact of Accommodation Strategies on English Language Learners' Test Performance. *Educational Measurements: Issues and Practice*, 19: 3, 16–26.
- Ahmed, A. and Pollitt, A. (2007) Improving the Quality of Contextualized Questions: and Experimental Investigation of Focus. *Assessment in Education: Principles, Policy and Practice*, 14: 2, 201–232.
- Anderson, R., Wilkinson, I. and Mason, M. (1987) *Do Errors on Classroom Reading Tasks Slow Growth in Reading?* Center for the Study of Reading Technical Report No. 404. Champaign: University of Illinois at Urbana-Champaign.
- Assessment Performance Unit (1985) *A Review of Monitoring in Mathematics 1978–82*. Slough: NFER.
- Bailey, A. (2000) Language Analysis of Standardized Achievement Tests: Considerations in the Assessment of English Language Learners. In Abedi, J., Bailey, A., Butler, F., Castellon-Wellington, M., Leon, S. and Mirocha, J. (eds) *The Validity of Administering Large-scale Content Assessments to English Language Learners: an Investigation from Three Perspectives (CSE Report 663)* 85–105. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Barshi, I. and Healy, A. (2002) The Effects of Mental Representation on Performance in a Navigation Task. *Memory and Cognition*, 30, 1189–1203.
- Beck, I., McKeown, M., Sinatra, G. and Loxterman, J. (1991) Revising Social Studies Text from a Text-processing Perspective: Evidence of Improved Comprehensibility. *Reading Research Quarterly*, 26, 251–276.

Bügel, K. and Buunk, B. (1996) Sex Differences in Foreign Language Text Comprehension: The Role of Interests Knowledge. *The Modern Language Journal*, 80: 1, 15–31.

Butler, F. and Stevens, R. (1997) *Accommodation Strategies for English Language Learners on Large-Scale Assessments: Student Characteristics and Other Considerations (CSE Technical Report 448)*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Chall, J. (1976) *The Great Debate: Ten Years Later, with a Modest Proposal for Reading Stages*. Paper presented at the Conference on Theory and Practice of Beginning Reading Instruction, University of Pittsburgh, Learning Research and Development Center, April 1976. Chall, J. S. and Conard, S. (1991) *Should Textbooks Challenge Students? The Case for Easier or Harder Books*. New York: Teachers College Press.

Chall, J. S. and Dale, E. (1995) *Readability revisited: the New Dale-Chall Readability Formula*. Cambridge, Mass.: Brookline Books.

Chan, D., Schmitt, N., DeShon, R., Clause, C. and Delbridge, K. (1997) Reactions to Cognitive Ability Tests: the Relationships between Race, Test Performance, Face validity Perceptions, and Test Taking Motivation. *Journal of Applied Psychology*, 82, 300–310.

Chapman, L.J. (1987) *Reading from 5–11 Years*. London: Open University Press.

Chi, M. T. H. (1985) Interactive Roles of Knowledge and Strategies in the Development of Organized Sorting and Recall. In Chipman, S. F. Segal, J. W. and Glaser, R. (eds), *Thinking and Learning Skills* (Vol. 2, pp. 457–484). Hillsdale, NJ: Erlbaum.

Clay, M. (1977) *Reading: The Patterning of Complex Behaviour*. Portsmouth, NH: Heinemann.

Cole, N. and Zieky, M. (2001) The New Faces of Fairness. *Journal of Educational Measurement*, 38: 4, 369–382.

Connor, U. (1996) *Contrastive Rhetoric: Cross-cultural Aspects on Second-language Writing*. Cambridge: Cambridge University Press.

Crisp, V. (2011) Exploring Features That Affect the Difficulty and Functioning of Science Exam Questions for Those with Reading Difficulties. *Irish Educational Studies*, 30: 3, 323–343.

Crisp, V. and Sweiry, E. (2006) Can a Picture Ruin a Thousand Words? The Effects of Visual Resources in Exam Questions. *Educational Research*, 48: 2: 139–154.

Dale, E. and Chall, J.S. (1948) A Formula for Predicting Readability. *Educational Research Bulletin*, 11–20, 27, 28, 37–54.

Department of Education and Science (1975) *A Language for Life (the Bullock Report)*. London: HMSO.

Dorans, N. J. and Kulick, E. M. (1983) *Assessing Unexpected Differential Item Performance of Female Candidates on SAT and TSWE Forms Administered in December 1977: an Application of the Standardization Approach (ETS Technical Report RR-83-9)*. Princeton, NJ: ETS.

Downing, J. and Leong, C. (1982) *Psychology of Reading*. New York: Macmillan.

Ecalte, J., Bouchafa, H., Potocki, A. and Magnan, A. (2011) Comprehension of Written Sentences as a Core Component of Children's Reading Comprehension. *Journal of Research in Reading*. Article first published online: 6 May 2011. Available at <http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9817.2011.01491.x/full> (accessed 28/8/12).

Edwards, R. and Gibbon, V. (1973) *Words Your Children Use: A Survey of the Words Used by Children in Infants' Schools with the Resultant Graded Vocabulary*. London: Burke.

Eyles, P., Skelly, J. and Lou Schmuck, M. (2003) Evaluating Patient Choice of Typeface Style and Font Size for Written Health Information in an Outpatient Setting. *Clinical Effectiveness in Nursing*, 7: 2, 94–98.

Fang, Z. (2006) The Language Demands of Science Reading in Middle School. *International Journal of Science Education*, 28: 5, 491–520.

Feely, M., Rubin, G. S., Ekstrom, K. and Perera, S. (2005) Investigation into Font Characteristics for Optimum Reading Fluency in Readers with Sight Problems. *International Congress Series*, 1282, 530–533.

Flesch, R. (1948) A New Readability Yardstick. *Journal of Applied Psychology*, 32, 221–233.

Fisher-Hoch, H., Hughes, S. and Bramley, T. (1997) *What Makes GCSE Exam Questions Difficult? Outcomes of Manipulating Difficulty of GCSE Questions*. Paper presented at the British Educational Research Association Annual Conference, September 11–14, University of York.

Francis, W. and Kucera, H. (1979) *Brown Corpus Manual*. Providence, Rhode Island: Brown University. Available at <http://icame.uib.no/brown/bcm.html> (accessed 16/8/12).

Fry, E. (1977) *Elementary Reading Instruction*. New York: McGraw-Hill.

Fulcher, G. (1989) Cohesion and Coherence in Theory and Reading Research. *Journal of Research in Reading*, 12: 2, 146–163.

Gambrell, L., Wilson, R. and Gantt, W. (1981) Classroom Observations of Task-attending Behaviors of Good and Poor Readers. *Journal of Educational Research*, 74, 400–404.

Gierl, M. (2005) Using Dimensionality-based DIF Analyses to Identify and Interpret Constructs That Elicit Group Differences. *Educational Measurement: Issues and Practice*, 24, 3–14.

Gierl, M., Bisanz, J., Bisanz, G. and Boughton, K. (2003) Identifying Content and Cognitive Skills That Produce Gender Differences in Mathematics: a Demonstration of the Multidimensionality-based DIF Analysis Paradigm. *Journal of Educational Measurement*, 40: 4, 281–306.

Gilliland, J. (1975) *Readability*. London: Hodder and Stoughton.

Graesser, A., McNamara, D. and Louwrese, M. (2003) What Do Readers Need to Learn in Order to Process Coherence Relations in Narrative and Expository Text. In Sweet, A. P. and Snow, C. E. (eds) *Rethinking Reading Comprehension* (pp. 82–98). New York: Guilford.

Gunning, R. (1952) *The Technique of Clear Writing*. New York, NY: McGraw-Hill International Book Co.

Guthrie, J. and Cox, K. (2001) Classroom Conditions for Motivation and Engagement in Reading. *Educational Psychology Review*, 13: 3, 283–302.

Guthrie, J. T., Solomon, A. and Rinehart, J. M. (1997) Literacy Issues in Focus: Engagement in Reading for Young Adolescents. *Journal of Adolescent and Adult Literacy*, 40: 6, 438–446.

Halliday, M., and Hasan, R. (1976) *Cohesion in English*. New York: Longman.

Hamilton, L. (1998) Gender Differences on High School Science Achievement Tests: Do Format and Content Matter? *Educational Evaluation and Policy Analysis*, 20: 3, 179–195.

Harris, T. L. and Hodges, R. (1995) *The Literacy Dictionary: the Vocabulary of Reading and Writing*. Newark, Del.: International Reading Association.

Harrison, C. (1984) *Readability in the classroom*. Cambridge: Cambridge University Press.

Hughes, L. E. and Wilkins, A. J. (2000) Typography in Children's Reading Schemes May Be Suboptimal: Evidence from Measures of Reading Rate. *Journal of Research in Reading*, 23: 3, 314–324.

Janan, D., Wray, D. and Pope, M. (2010) Paradigms in Readability Research. *International Journal of Arts and Sciences*, 3: 17, 19–29.

Johnson, K. (1998) *Readability: Measuring the Reading Age of Books and Other Reading Matter*. Available at [www.timetabler.com/reading.html](http://www.timetabler.com/reading.html) (accessed 16/8/12).

Johnston, P. H. (1984) Prior Knowledge and Reading Comprehension Test Bias. *Reading Research Quarterly*, 19: 3, 219–239.

Just, M. and Carpenter, P. A. (1987) *The Psychology of Reading and Language Comprehension*. Newton, MA: Allyn and Bacon.

Just, M. A., Carpenter, P. A. and Wooley, J. D. (1982) Paradigms and Processes in Reading Comprehension. *Journal of Experimental Psychology*, 111: 2, 228–238.

Kieffer, M., Lesaux, N., Rivera, M. and Francis, D. (2009) Accommodations for English Language Learners Taking Large-scale Assessments: a Meta-analysis on Effectiveness and Validity. *Review of Educational Research*, 79: 3, 1168–1201.

Kintsch, Walter (1974) *The Representation of Meaning in Memory*. Hillsdale: Lawrence Erlbaum.

Kintsch, W. and Keenan, J. (1973) Reading Rate and Retention as a Function of the Number of Propositions in the Base Structure of Sentences. *Cognitive Psychology*, 5: 257–274.

Kletzien, S. (1991) Strategy Use by Good and Poor Comprehenders Reading Expository Text of Differing Levels. *Reading Research Quarterly*, 26, 67–86.

Louwerse, M. M. and Graesser, A. C. (2004) Coherence in Discourse. In Strazny, P. (ed.) *Encyclopedia of Linguistics*. Chicago: Fitzroy Dearborn.

Lund, O. (1999) *Knowledge Construction in Typography: The Case of Legibility Research and the Legibility of Sans Serif Typefaces*. Thesis submitted for the degree of Doctor of Philosophy. Reading: The University of Reading, Department of Typography and Graphic Communication.

McLaughlin, G. Harry (May, 1969) *SMOG Grading — a New Readability Formula*. *Journal of Reading*, 12 (8), 639–646.

McNally, J. and Murray, W. (1968) *Key Words to Literacy and the Teaching of Reading: a Basic Word List for Developing Early Reading and Writing Skills*. London, Schoolmaster Publishing.

Maccoby, E. (1990) Gender and Relationships: a Developmental Account. *American Psychologist*, 45: 4, 513–20.

Millard, E. (1997) *Differently Literate: Boys, Girls and the Schooling of Literacy*. London: Falmer Press.

Milone, M. (2008) *The Development of ATOS: The Renaissance Readability Formula*. Wisconsin Rapids, WI: Renaissance Learning.

Morgan, P., Farkas, G. and Hibel, J. (2008) Matthew Effects for Whom? *Learning Disability Quarterly*, 31: 4, 187–198.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J. and Kennedy, A. M. (2003) *PIRLS 2001 International Report: IEA's Study of Reading Literacy Achievement in Primary School*. Chestnut Hill, MA: Boston College.

Nation, P. and Waring, R. (1997) Vocabulary Size, Text Coverage and Word Lists. In Schmitt, N. and McCarthy, M. (eds) *Vocabulary: Description, Acquisition and Pedagogy* (6–19). Cambridge: Cambridge University Press. Available at [www.lexutor.ca/research/nation\\_waring\\_97.html](http://www.lexutor.ca/research/nation_waring_97.html) (accessed 16/8/12).

Newbold, N. and Gillam, L. (2010) The Linguistics of Readability: The Next Step for Word Processing. *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics and Writing* (June), 65–72. Available at <http://www.aclweb.org/anthology/W/W10/W10-0409.pdf> (accessed 28/8/12).

Nickson, M. and Green, S. (1996) *A Study of the Effects of Context in the Assessment of the Mathematical Learning of 10/11 Year Olds*. Paper presented to the 1996 British Educational Research Association. Available at [www.cambridgeassessment.org.uk/ca/](http://www.cambridgeassessment.org.uk/ca/) (accessed 16/8/12).

O'Reilly, T. and McNamara, D. (2007) Reversing the Reverse Cohesion Effect: Good Texts Can Be Better for Strategic, High-knowledge Readers. *Discourse Processes*, 43, 121–152.

Ofqual (2010) *Fair Access by Design*. Coventry: Ofqual.

Ofqual (2011) *Regulatory Framework for National Assessments: National Curriculum and Early Years Foundation Stage*. Coventry: Ofqual.

Ofqual (2012) *Guidance on the Principles of Language Accessibility in National Curriculum Assessments*. Coventry: Ofqual.

Ofqual (2012) *Guidance on Monitoring Access to National Curriculum Assessments*. Coventry: Ofqual.

Ofqual (2012) *Monitoring Access to National Curriculum Assessments: Research Background*. Coventry: Ofqual.

Ozuru, Y., Dempsey, K. and McNamara, D. (2009) Prior Knowledge, Reading Skill, and Text Cohesion in the Comprehension of Science Texts. *Learning and Instruction*, 19, 228–242.

Perera, K. (1980) The Assessment of Linguistic Difficulty in Reading Material, *Educational Review*, 32: 2, 151–161.

Pillai, K. G., Katsikeas, C. S. and Presib, C. (2012) Print Advertising: Type Size Effects. *Journal of Business Research*. Volume 65, Issue 6 June 2012, pages 865-868.

Pollitt, A., Hutchinson, C., Entwistle, N. and de Luca, C. (1985) *What Makes Examination Questions Difficult?* Edinburgh: Scottish Academic Press.

Rasinski, T. V. (1990) *The Effects of Cued Phrase Boundaries on Reading Performance: a Review*. Kent, Ohio: Kent State University, Department of Teacher Development and Curriculum Studies.

Reid, J. (1972) Children's Comprehension of Syntactic Features Found in Extension Readers. In Reid, J. (ed) *Reading: Problems and Practices*. London: Ward Lock.

Robinson, J. (2010) The Effects of Test Translation on Young English Learners' Mathematics Performance. *Educational Researcher*, 39: 8, 582–590.

Roussos, L. and Stout, W. (1996) A Multidimensionality-based DIF Analysis Paradigm. *Applied Psychological Measurement*, 20, 355–371.

Saarnio, D. A., Oka, E. R. and Paris, S. G. (1990) Developmental Predictors of Children's Reading Comprehension. In Carr, T. H. and Levy, B. A. (eds), *Reading and Its Development: Components Skills Approaches* (57–79). San Diego, CA: Academic Press.

Schagen, I. and Sainsbury, M. (1996) Multilevel Analysis of the Key Stage 1 National Curriculum Assessment Data in 1995. *Oxford Review of Education*, 22: 3, 265–272.

Schmit, M. and Ryan, A. (1992) Test-taking Dispositions: a Missing Link? *Journal of Applied Psychology*, 77, 629–637.

Senter, R.J. and Smith, E. A. (November, 1967) Automated Readability Index. Wright-Patterson Air Force Base. p. iii. AMRL-TR-6620. Available online at [www.dtic.mil/cgi-bin/GetTRDoc?AD=AD0667273](http://www.dtic.mil/cgi-bin/GetTRDoc?AD=AD0667273) (accessed 28/8/12).

Shorrocks-Taylor, D., Curry, J., Swinnerton, B. and Nelson, N. (2003) National Curriculum Mathematics Tests in England at Key Stage 2: Weights and Measures? *Oxford Review of Education*, 29: 1, 51–66.

Shuard, H. and Rothery, A. (1984) *Children Reading Mathematics*. London: John Murray.

Sonnleitner, P. (2008) Using the LLTM to Evaluate an Item Generating System for Reading Comprehension. *Psychology Science Quarterly*, 50, 345–362.

Spache, G. (1953) A New Readability Formula for Primary-Grade Reading Materials. *The Elementary School Journal*, 53 (7), 410–413.

Spache, G. (1974) *Good Reading for Poor Readers*. Champaign, IL: Garrard.

Stanovich, K. (1986) Matthew Effects in Reading: Some Consequences of Individual Differences in the Acquisition of Literacy. *Reading Research Quarterly*, 21: 4, 360–407.

Stone, E., Cook, L., Laitusis, C. C. and Cline, F. (2010) Using Differential Item Functioning to Investigate the Impact of Testing Accommodations on an English-Language Arts Assessment for Students Who Are Blind or Visually Impaired. *Applied Measurement in Education*, 23: 2, 132–152.

Stricker, L. and Emmerich, W. (1999) Possible Determinants of Differential Item Functioning: Familiarity, Interest, and Emotional Reaction. *Journal of Educational Measurement*, 36: 4, 347–366.

Thompson, C., and Shapiro, L. (2007) Complexity in Treatment of Syntactic Deficits. *American Journal of Speech-Language Pathology*, 18, 30–42.

Thompson, S. J., Thurlow, M. L. and Malouf, D. (2004) Creating Better Tests for Everyone Through Universally Designed Assessments. *Journal of Applied Testing Technology*. 6: 1. Available at <http://www.testpublishers.org/assets/documents/volume%206%20issue%201%20Creating%20%20better%20tests.pdf> (accessed 28/8/12).

Tinker, M. A. (1963) *Legibility of Print*. Iowa: Iowa State University Press.

Waller, R. H. W. (1991) *Designing Reports and Presentations*. Redhill: Monotype Desktop Solutions.

Wellington, J. and Osborne, J. (2001) *Language and Literacy in Science Education*. Buckingham: Open University Press.

White, A. (2004) *Thinking in Type: The Practical Philosophy of Typography*. New York: Allworth Press.

Wilkins, A., Cleave, R., Grayson, N. and Wilson, L. (2009) Typography for Children May Be Inappropriately Designed. *Journal of Research in Reading*, 32: 4, 402–412.

Zumbo, B. (2007) Three Generations of DIF Analyses: Considering Where It Has Been, Where It Is Now, and Where It Is Going. *Language Assessment Quarterly*, 4: 2, 223–233.

## Appendix: The 1,000 most commonly used words in English text

(Taken from the General Service List, available at <http://jbauman.com/aboutgsl.html>)

a, ability, able, about, above, accept, accord, account, across, act, action, active, actual, add, address, admit, advance, advantage, affair, after, afternoon, again, against, age, agency, ago, agree, ahead, air, all, allow, almost, alone, along, already, also, although, always, among, amount, and, animal, another, answer, any, anyone, anything, appear, apply, approve, argue, arm, army, around, arrange, arrive, art, article, as, ask, association, at, attack, attempt, attend, attention, audience, average, avoid, away
back, bad, balance, ball, bank, bar, base, basic, basis, battle, be, bear, beat, beauty, because, become, bed, before, begin, behaviour, behind, believe, below, best, better, between, beyond, big, bill, bit, black, block, blood, blue, board, boat, body, book, both, bottle, bottom, boy, break, bridge, bright, bring, broad, brother, build, burn, business, but, buy, by
call, camp, can, capital, captain, car, care, carry, case, catch, cattle, cause, cent, centre, century, certain, chance, change, character, charge, check, chief, child, choice, choose, church, circle, citizen, city, claim, class, clean, clear, close, clothe, club, coat, cold, college, colour, combine, come, comfort, command, committee, common, company, compare, complete, compose, concern, condition, conscious, consider, contain, continue, control, cool, corner, cost, could, council, count, country, course, court, cover, critic, cross, crowd, cry, current, cut
daily, dance, danger, dark, date, day, dead, deal, death, decide, decision, declare, deep, defence, degree, demand, department, depend, dependent, describe, desire, destroy, detail, determine, develop, die, difference, different, difficult, difficulty, dinner, direct, direction, director, discover, discuss, discussion, distance, district, division, do, doctor, dog, dollar, door, doubt, down, draw, dream, dress, drink, drive, drop, dry, due, during, dust, duty
each, early, earth, east, easy, eat, edge, education, effect, effective, effort, either, election, electric, else, employ, employee, encourage, end, enemy, engineer, English, enjoy, enough, enter, entire, equal, escape, especially, essential, even, evening, event, ever, every, everyone, everything, exact, examine, example, except, excite, exercise, exist, existence, expect, expense, experience, experiment, explain, express, extend, extent, extreme, eye
face, fact, fail, fair, faith, fall, fame, familiar, family, far, farm, fast, father, favour, fear, feed, feel, few, field, fight, figure, fill, film, find, fine, finger, finish, fire, firm, first, fit, fix, floor, flow, fly, follow, food, foot, for, force, foreign, forget, form, former, forward, frame, free, freedom, frequent, friend, from, front, full, further, future
gain, game, garden, gas, general, get, girl, give, glass, go, god, good, govern, governor, great, green, ground, group, grow, growth, guest, gun

hair, half, hall, hand, handle, hang, happen, happy, hard, hardly, have, he, head, health, hear, heart, heat, heavy, help, here, high, hill, history, hit, hold, hole, home, honour, hope, horse, hospital, hot, hotel, hour, house, how, however, human, husband
I, idea, ideal, if, imagine, immediate, importance, important, improve, in, inch, include, increase, indeed, industry, influence, inform, inside, instead, interest, international, into, island, it
join, judge, just, justice
keep, kill, kind, king, kitchen, know, knowledge
lack, lady, land, language, large, last, late, latter, laugh, law, lay, lead, learn, least, leave, left, leg, length, less, let, letter, level, lie, life, light, like, likely, limit, line, list, listen, literature, little, live, load, local, lock, long, look, lose, loss, lot, love, low
machine, main, make, man, manage, manner, manufacture, many, mark, market, marriage, marry, mass, master, match, material, matter, may, maybe, mean, measure, medical, meet, member, memory, mention, mere, middle, might, mile, mind, minute, miss, model, modern, moment, money, month, moral, more, morning, most, mother, motor, mountain, mouth, move, much, murder, music, must
name, nation, nature, near, necessary, need, neither, never, new, news, newspaper, next, night, no, none, nor, north, not, note, nothing, notice, now, number
object, observe, occasion, of, off, offer, office, officer, official, often, oil, old, on, once, one, only, open, operate, operation, opinion, opportunity, or, order, organize, origin, other, out, outside, over, own
page, pain, paint, paper, parent, park, part, particular, party, pass, past, patient, pattern, pay, peace, people, per, perfect, perform, performance, perhaps, permit, person, pick, picture, piece, place, plan, plant, play, please, poem, poet, point, police, political, pool, poor, popular, population, position, possible, post, pound, power, practical, practice, prepare, present, president, press, pressure, pretty, prevent, price, private, probable, problem, produce, product, production, profession, program, progress, promise, proper, property, propose, prove, provide, public, pull, pure, purpose, push, put
quality, question, quick, quiet, quite
race, radio, raise, rapid, rate, rather, reach, read, ready, real, realize, reason, reasonable, receive, recent, recognize, recommend, record, red, reduce, refer, reflect, refuse, regard, regular, relate, relation, relative, religion, remain, remark, remember, repeat, replace, reply, report, represent, representative, respect, responsible, rest, result, return, ride, right, rise, river, road, rock, roll, room, round, rule, run
sale, same, sample, save, say, scene, school, science, sea, search, season, seat, second, secret, secretary, see, seem, sell, send, sense, separate, serious, serve, service, set, settle, several, shake, shall, shape, share, sharp, she, shelter, ship, shoot, shop, short, should, shoulder, show, side, sight, sign, signal, simple, since, sing, single, sit, situation, size, skill, sleep, slight, slow, small, smile, so, social,

society, soft, soldier, solid, some, someone, something, sometimes, son, song, soon, sort, sound, south, space, speak, special, speed, spend, spirit, spot, spread, spring, square, staff, stage, stand, standard, start, state, station, stay, step, stick, still, stock, stop, store, story, straight, strange, street, strength, strike, strong, student, study, subject, success, such, sudden, suffer, suggest, suit, summer, sun, supply, support, suppose, sure, surface, surprise, system

table, take, talk, tax, teach, telephone, tell, temperature, tend, term, test, than, that, the, then, there, therefore, these, they, thick, thin, thing, think, this, those, though, through, throw, thus, time, title, to, today, together, too, tooth, top, total, touch, toward, town, trade, train, travel, treat, tree, trial, trip, trouble, true, truth, try, turn, type

under, understand, union, unit, unite, university, unless, until, up, upon, use, usual  
value, various, very, view, visit, voice, vote

wage, wait, walk, wall, want, war, warm, wash, watch, water, wave, way, we, weak, weapon, wear, week, weight, well, west, western, what, whatever, when, where, whether, which, while, white, who, whole, why, wide, wife, will, win, wind, window, wine, wish, with, within, without, woman, wonder, word, work, world, worry, worth, would, write, wrong

yard, year, yes, yet, you, young, youth

We wish to make our publications widely accessible. Please contact us if you have any specific accessibility requirements.

First published by the Office of Qualifications and Examinations Regulation in 2012

© Crown copyright 2012

You may re-use this publication (not including logos) free of charge in any format or medium, under the terms of the [Open Government Licence](#). To view this licence, visit [The National Archives](#); or write to the Information Policy Team, The National Archives, Kew, Richmond, Surrey, TW9 4DU; or email: [psi@nationalarchives.gsi.gov.uk](mailto:psi@nationalarchives.gsi.gov.uk)

This publication is also available on our website at [www.ofqual.gov.uk](http://www.ofqual.gov.uk)

Any enquiries regarding this publication should be sent to us at: Office of Qualifications and Examinations Regulation

Spring Place  
Coventry Business Park  
Herald Avenue  
Coventry CV5 6UB

2nd Floor  
Glendinning House  
6 Murray Street  
Belfast BT1 6DN

Telephone 0300 303 3344

Textphone 0300 303 3345

Helpline 0300 303 3346