REVIEW

June 2004

The effect of grammar teaching (syntax) in English on 5 to 16 year olds' accuracy and quality in written composition

Review conducted by the English Review Group

NAME OF GROUP AND INSTITUTIONAL LOCATION

EPPI Review Group for English
Department of Educational Studies, University of York, UK

AUTHORS AND REVIEW TEAM

Richard Andrews, Department of Educational Studies, University of York
Sue Beverton, School of Education, University of Durham
Terry Locke, Arts and Language Education Department, University of Waikato, New Zealand
Graham Low, Department of Educational Studies, University of York
Alison Robinson, Department of Educational Studies, University of York
Carole Torgerson, Department of Educational Studies, University of York
Die Zhu, Department of Educational Studies, University of York

ADVISORY GROUP MEMBERSHIP

Judith Bennett, Department of Educational Studies, University of York
James Durran, Parkside Community College, Cambridge
Polly Griffith, Chair of Governors, Millthorpe School, York
Nick McGuinn, Department of Educational Studies, University of York
Gloria Reid, Kingston-upon-Hull Local Education Authority
Peter Taylor, Oaklands School and All Saints School, York
Ian Watt, Department of Health Sciences, University of York

Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) support
Diana Elbourne
Jo Garcia

ACKNOWLEDGEMENTS AND CONFLICTS OF INTEREST

LIST OF ABBREVIATIONS

CT   Controlled trial
DES   Department of Education and Science (England and Wales)
DfEE   Department for Education and Employment (England and Wales)
PGCE   Postgraduate Certificate in Education
QA   Quality assurance
QCA   Qualifications and Curriculum Authority (England and Wales)

GLOSSARY

Accuracy of writing
Accuracy in terms of sentence structure and correct use of punctuation with standard written English.

Coherence
Relationships that link sentences together to form a meaningful flow of ideas or propositions. The links between sentences are often inferred, rather than explicitly flagged.

Cohesion
Grammatical or lexical (word-level) relationships that bind different parts of a text together: for example, 'however', 'on the one hand…', 'on the other hand…'.

Contextualised grammar teaching
Grammar teaching that takes account of the function of sentences and texts in context, and also of the relationship of sentences to higher (e.g. text) and lower (e.g. phrase, clause, word, morpheme ['the smallest meaningful unit of grammar']) units of language description.

De-contextualised grammar teaching
Sometimes known as 'traditional' grammar teaching, this focuses on the internal dynamics and structure of the sentence or text, not in the context of written production (e.g. drill and practice).

Deep syntactic structures
These are the projected abstract underlying structures of a sentence (as opposed to surface structures); more loosely, deep and surface structures form a binary contrasting pair of descriptors, the first being the supposed underlying meaning, and the second the actual sentence we see or hear.

'Functional' grammar
The term used to describe Halliday's systemic-functional grammar (Halliday and Hasan, 1985). Such a grammar goes beyond the description or prescription or generation of sentences or texts. It aims to relate text and sentence to context and meaning.

Language awareness
An approach to teaching about language that aims to raise awareness of different aspects of language, as opposed to formal grammar teaching.

ii

Learning difficulties
General difficulties with learning, often assumed to face about 20% of the school population from time to time.

Meta-language
A diction (specialised subset of language) used to discuss language, e.g. 'noun', 'syntax'.

Oracy
The spoken equivalent of 'literacy'. The term is derived from an analogy with 'literacy'.

Paradigmatic
A set of linguistic items in which any member of the set can be substituted (grammatically) for another member. Paradigmatic items are in an 'or' relationship, whereas syntagmatic items (their opposite) are in an 'and' relationship to each other. For example, nouns and verbs each form a paradigmatic class.

Paragraph composition
Paragraphs have no grammatical status as such, but their arrangement within a text (e.g. 'the five-paragraph essay' in the US tradition), is considered part of teaching textual grammar.

'Pedagogic' grammar
The distillation (usually of a traditional grammar) as used in textbooks for first or second language teaching.

Punctuation
Surface markers for sentence structure, or, in the case of exclamation marks and question marks, indicators of tone and function.

Quality of writing
Quality in terms of a set of criteria: for example, 'cohesion', 'imaginativeness', 'appropriateness of style', 'verve'. Usually judged inter-subjectively by a panel of experts (e.g. teachers).

Sentence-combining
A teaching technique for linking sentences *horizontally*, i.e. not via their meaning or sub-grammatical character, but with connectives (e.g. conjunctions) or syntagmatically (see 'syntagmatic'). It can also cover sentence-embedding and other techniques for expanding and complicating the structure of sentences.

Sentence-diagramming
A technique deriving from structural and transformational grammars in which relationships between parts of a sentence are presented diagrammatically, often in tree-diagram form.

'Sentence' level grammar teaching
Teaching about the structural rules of sentence creation.

Specific learning difficulties
Dyslexia and other specific difficulties with language learning.

Syntagmatic
See 'paradigmatic'. Syntagmatic relationships can be conceived as in a chain or sequence, for example, the relationship between nouns and verbs in a sentence.

Syntax
Constraints which control acceptable word order within a sentence, or dominance relations (like head noun + relative clause).

'Text' level grammar teaching
Teaching about the cohesion* of a stretch of written composition. The term 'text grammar' applies the notion of grammar to whole texts, with an assumption of semantic (meaning), or pragmatic (meaning in use) coherence*.
See above

Text structure
Rules governing the internal arrangement of whole texts.

Traditional grammar
Sentence grammars that tend to focus on the internal elements of the sentence, classifying 'parts of speech' and describing (and sometimes prescribing) the relationship between parts of speech.

Transformative/generative grammar
A transformative grammar attempts to systematise the changes that take place between the deep structures in language patterning and surface structures (i.e. the actual utterances made by speakers and writers); such a grammar is termed 'generative' because it is thought to be able to generate sentences or meaningful utterances, as opposed to merely describing or prescribing rules for their information.

Written composition
'Composition' is the term used to describe the putting together of words in an extended piece of writing.

TABLE OF CONTENTS

SUMMARY

Background

A systematic review is needed in order to ask the question: What is the effect of grammar teaching on the accuracy and quality of 5 to 16 year-olds' written composition?

This perennial question has haunted the teaching of English for over a century. Although there have been extensive reviews of the question, views remain polarised, with a belief among some teachers, newspaper editors and members of the public, that such teaching is effective, and among others that it is ineffective. A systematic review is therefore required to provide an authoritative account of the results of research into the question.

The objectives of the review are as follows:

- to map the field of research on the effects of text- and sentence-level grammar teaching on writing in English-speaking countries for pupils aged between 5 and 16

- to undertake two distinct but complementary in-depth reviews in the field of sentence-level grammar: the effect of teaching syntax on accuracy and quality in written composition (in 2003-4); the effect of teaching sentence-combining on accuracy and quality in written composition (in 2004-5)

The present review concerns the effect of teaching syntax on the accuracy and quality of written composition.

One previous systematic review has been published in the broader field of the effect of grammar teaching on written composition. In 1986, Hillocks published a meta-analysis of experimental studies designed to improve the teaching of written composition. He analysed the experimental research between 1960 and 1982 and concluded that grammar instruction led to a statistically significant decline in student writing ability, the only instructional method of those examined not to produce gains in writing ability.

Methods used in the review

Systematic review methods were used throughout this review, using the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) guidelines and tools for conducting a systematic review (EPPI-Centre, 2002a, 2002b and 2002c).

Studies were included in the systematic map if they looked at the effect of grammar teaching in English on 5 to 16 year olds' accuracy and quality in written composition. The criteria for including and excluding studies for the in-depth review on the effect of teaching 'syntax' were refined after the systematic map was drawn.

Reports were identified from the following sources:
- searching of electronic bibliographic databases: Educational Resources Information Center (ERIC); PsycINFO; and Social Science Citation Index (SSCI)
- citations in reference lists of all included systematic and non-systematic reviews
- personal contacts

We applied the inclusion and exclusion criteria successively to the titles and abstracts and the full reports with quality assurance (QA) screening supplied by the EPPI-Centre.

The studies remaining after application of the criteria were keyworded using the EPPI-Centre's *Core Keywording Strategy* (EPPI-Centre, 2002a) and online database software, *EPPI-Reviewer* (EPPI-Centre, 2002b). Additional review-specific keywords which are specific to the context of the review were added to those of the EPPI-Centre. Again, QA was provided by the EPPI-Centre.

Studies identified as meeting the inclusion criteria for the in-depth review were analysed in depth using the EPPI-Centre's detailed *Data-Extraction Guidelines* (EPPI-Centre, 2002c), together with its online software, *EPPI-Reviewer*® (EPPI-Centre, 2002b). Three components were identified to help in making explicit the process of apportioning different weights to the findings and conclusions of different studies. Such weights of evidence are based on the following:

(A) the soundness of studies (internal methodological coherence), based upon the study only
(B) the appropriateness of the research design and analysis used for answering the review question
(C) the relevance of the study topic focus (from the sample, measures, scenario, or other indicator of the focus of the study) to the review question
(D) an overall weight taking into account (A), (B) and (C)

The data were then synthesised to bring together the studies which answer the review question and which meet the quality criteria relating to appropriateness and methodology. A narrative synthesis was undertaken. It was not felt to be appropriate to conduct a statistical meta-analysis.

Data-extraction and assessment of the weight of evidence brought by the study to address the review question was conducted by pairs of Review Group members, working first independently and then comparing their decisions before coming to a consensus. Members of the EPPI-Centre helped in data-extraction and quality appraisal of a sample of studies.

Identifying and describing studies: results

A total of 4,566 potentially relevant papers were identified from the initial searches. After screening for relevance to the review using the pre-established inclusion and exclusion criteria, 58 papers were included in the systematic map of research in the field. The 58 papers comprised 25 papers containing 24 systematic and non-systematic reviews, and 33 papers containing 31 primary studies. All the included primary studies were study type C, i.e. evaluations: 30

researcher-manipulated evaluations and one naturally-occurring evaluation. Of the 30 researcher-manipulated evaluations, seven were randomised controlled trials (RCTs), 13 were controlled trials (CTs), eight were pre- and post-test studies, and two were evaluations of 'other' designs.

Sixteen out of the 24 reviews explored the teaching of 'syntax'. Of these 16, 12 provided a conclusion about the effect of syntax teaching on the accuracy and quality of pupils' writing. None of these 12 reviews of the teaching of syntax concluded that teaching traditional or transformative/generative grammar had a positive effect on the quality and accuracy of 5 to 16 year-olds' written compositions. The results of these reviews provide the context for our discussion of the results of our review.

Of the 28 studies that reported on sentence-level grammar teaching, 20 focused on sentence-combining and 10 focused on other aspects of syntax (the focus of the in-depth review). A much smaller proportion focused on punctuation (n = 3), and only one study focused on sentence-diagramming. Three studies investigated the teaching of both sentence-combining and syntax. One study focused on sentence-combining and punctuation; one on syntax, punctuation and sentence-diagramming; and one on punctuation alone.

In-depth review: results

Ten studies were identified for the in-depth review. These studies were identified through the application of the review-specific keyword 'syntax' to the primary studies in the map.

The ten studies selected for in-depth review were all researcher-manipulated experimental studies, of which two were randomised controlled trials (Bateman and Zidonis, 1966; Fogel and Ehri, 2000); two were controlled trials (Elley *et al.*, 1975, 1979; Stock, 1980); four used pre- and post-tests (Hilfman, 1970; McNeill, 1994; Roberts and Boggase, 1992; Rousseau and Poulson, 1985); one was a curriculum evaluation (Satterfield and Powers, 1996); and one a single subject ABACA design (Stone and Serwatka, 1982).

The narrative overview must begin with the studies rated high and high/medium or medium/high. These are Elley *et al.* (1975, 1979) (high to medium); Bateman and Zidonis (1966) (medium to high); and Fogel and Ehri (2000) (high).

It is not possible to synthesise systematically the results of the Elley *et al.* and Bateman and Zidonis studies. First, the transformational grammatical approach of Elley *et al.*, based as it is on materials from the Oregon Curriculum (Kitzhaber, 1968), uses – we assume – different intervention materials from the unspecified 'special grammatical materials' of Bateman and Zidonis. Second, the analytical framework of the two studies is different, with Elley *et al.* using 12 variables for analysis and Bateman and Zidonis, 46. Third, we cannot rule out from either study, for different reasons, methodological invalidity or unreliability. Fourth, there is insufficient detail given in Bateman and Zidonis of the intervention or of the analytical tools used (hence the lower rating than Elley *et al.* in terms of weight of evidence). Fifth, there is no clear comparability between the two studies because Elley *et al.* use what they call a 'transformational' approach, and Bateman and Zidonis use a 'generative' approach to transformational/generative

grammar. The relationship between the two, and to transformational and generative grammars and theories, is not clearly articulated.

In summary, Elley *et al.* conclude that syntax teaching, whether traditional or transformational, has virtually no influence on the language growth of typical secondary school students. Bateman and Zidonis conclude, tentatively, that a generative grammar approach does make a difference to syntactic quality and to the control of malformed sentences. Because of the relative quality of the two studies, methodologically, the results of the Elley *et al.* study have a higher weight of evidence. However, neither study can be said to be conclusive. Fogel and Ehri present a different kind of study in which mastery of standard English written forms is improved for elementary school African-American pupils by a process of exposure, strategies for labelling and identifying grammatical features and, crucially, practising writing in these forms and receiving teacher feedback. However, short-term feedback is not enough to cause change in pupils of this age. As the authors point out,

> further research is needed to determine whether more extensive and repeated use of the procedures would result in increased achievement; [because] instruction was limited to six forms…it is not clear whether findings would generalize to other more complex syntactic forms [nor] whether the performance differences that were observed would be maintained over time. These remain questions for further research (Fogel and Ehri, 2000, p 230).

Findings and implications

The results of the present in-depth review point to one clear conclusion: that there is no high quality evidence to counter the prevailing belief that the teaching of the principles underlying and informing word order or 'syntax' has virtually no influence on the writing quality or accuracy of 5 to 16 year-olds. This conclusion remains the case whether the syntax teaching is based on the 'traditional' approach of emphasising word order and parts of speech, or on the 'transformational' approach, which is based on generative-transformational grammar.

Nearly all our included studies were experimental (i.e. researcher-manipulated as opposed to naturally-occurring evaluations), a highly appropriate design for testing causality.

In terms of practice, the main implication of our findings is that there is no high quality evidence that the teaching of grammar, whether traditional or generative/transformational, is worth the time if the aim is the improvement of the quality and/or accuracy of written composition. This is not to say that the teaching of such grammar might not be of value in itself, or that it might lead to enhanced knowledge and awareness of how language works, and of systems of language use. But the clear implication, based on the available high quality research evidence, is that the evidence base to justify the teaching of grammar in English to 5 to 16 year-olds in order to improve writing is very small.

It was not our brief in the present review to suggest what does work in improving the quality and accuracy of writing in English for 5 to 16 year-olds, but the

implication is that, if there is little evidence that formal grammar teaching of syntax works, then practices based on theories such as 'you learn to write by writing' need to be given more credence and subject themselves to further systematic review. Whether there is space in the curriculum to teach syntax for its own sake, or for other purposes, remains to be seen.

The implications for further research are various. Despite a hundred years of concern about the issue of the teaching of grammar and thousands of research studies, the high quality research base for claiming the efficacy of syntax teaching is small. The first implication, then, is that there should be a conclusive, large-scale and well-designed randomised controlled trial to answer the question about whether syntax teaching does improve the writing quality and accuracy of 5 to 16 year-olds. Such a study should have a longitudinal dimension to test whether any significant effects are sustained.

While we do not claim the final word on the question, the present review has been the largest systematic review in the history of research on the topic to date. This does not mean that other reviews of different aspects of the question of the relationship between grammar teaching and writing quality and accuracy cannot be undertaken. The specific focus of this review has been on the teaching of syntax and a complementary review we are undertaking is on sentence-combining, both of which come under the umbrella of 'grammar' teaching.

There are limitations to this particular review. The teaching of traditional grammar or syntax to improve written composition tends to ignore the levels of language immediately below and above the sentence; morphological structures in language below the level of the sentence; and paragraph and textual levels above the level of the sentence.

Despite the above reservation, we hope to have established a landmark in studies on the effectiveness of syntax teaching in the development of writing quality and accuracy in school-age children. If this is a landmark, it points the way to further research in the field, where the territory of debate will be somewhat different. We now know that there is no high quality evidence that teaching of traditional grammar or syntax (or the direct teaching of formal or generative/transformational grammars) is effective with regard to writing development. Having established that much, we can now go on to research what is effective, and to ask clearer and more pertinent questions about what works in the development of young people's literacy.

1. BACKGROUND

1.1 Aims and rationale for current review

A systematic review is needed in order to ask the question: What is the effect of grammar teaching on the accuracy and quality of 5 to 16 year-olds' written composition?

This perennial question has haunted the teaching of English for over a century. Although there have been extensive reviews of the question (e.g. Macaulay, 1947; Wilkinson, 1971; Wyse, 2001), views remain polarised, with a belief among some teachers, newspaper editors and members of the public, that such teaching is effective and among others that it is ineffective. A systematic review is therefore required to provide an authoritative account of the results of research into the question.

The aim of the review is to shed conclusive light on the effect (or otherwise) of grammar teaching on writing by 5 to 16 year-olds in English.

The objectives are as follows:

- to map the field of research on the effects of text- and sentence-level grammar teaching on writing in English-speaking countries for pupils aged between 5 and 16

- to undertake two distinct but complementary in-depth reviews in the field of sentence-level grammar: the effect of teaching syntax on accuracy and quality in written composition (in 2003-4); the effect of teaching sentence-combining on accuracy and quality in written composition (in 2004-5)

Research question for systematic map

What is the effect of grammar teaching in English on 5 to 16 year olds' accuracy and quality in written composition?

Research questions for in-depth reviews

- ***What is the effect of teaching syntax in English on 5 to 16 year olds' accuracy and quality in written composition?*** (the current review)

- ***What is the effect of teaching sentence-combining in English on 5 to 16 year-olds' accuracy and quality in written composition?*** (the complementary review)

1.2 Definitional and conceptual issues

A very short history of grammar teaching: understanding the research context

We can divide the understanding of the nature of grammar, its place within language learning and the teaching of grammar, into broad phases. Hudson (1992) suggests two phases to the understanding and teaching of formal written grammars.

According to Hudson, the first phase runs from 300 BC to 1957. This broad sweep of the history of grammars and grammar teaching has as its common strand the description of language and the subsequent prescription in 'grammar textbooks' in terms of how to write. The basic approach of these grammars is paradigmatic: that is, classes and categories of the language were defined, and these were then taught as a means to write the language. In the Renaissance, the principle of a scientific classificatory approach to written language gave rise to Grammar in the curriculum (the other disciplines were Rhetoric and Logic, precursors to discourse analysis, mathematics and philosophy) and, in turn, to grammar schools. Grammar was often taught in this period via *progymnasmata*, or exercises based on exemplary models of textual and sentence structure.

The publication of Chomsky's *Syntactic Structures* (1957) marks the beginning of the second of these phases. His approach is more syntagmatic than paradigmatic. That is to say, it takes a structuralist approach, assuming that language can be described cross-sectionally or at any one moment in history in terms of a coherent system of rules. Such an approach is part of the tradition of cognitive neuro-scientific theories of language production in that it is interested in the structural relationships between words, phrases and clauses in sentences, rather than in classificatory categories or 'parts of speech'. Chomsky's theory, with its distinction between deep syntactic structures and surface manifestations in speech and in writing, gave rise to generative and transformative grammars (see Damasio, 2000; Pinker, 1995). These grammars operated from basic principles in the construction of meaning that Chomsky claims existed as universals in all languages and which were intended to be able to generate intelligible sentences. Such generative capacity involved a *transformation* from deep structural rules and formulae to the actual utterances of everyday speech and writing.

At around the same time in the UK, Halliday was starting to construct what later became known as 'Functional Grammar'. One of Halliday's main contributions to the understanding of how language works was to combine the paradigmatic and syntagmatic, building on Firth's (1935, p 27) idea of the need to see formal and substantive aspects of language operating purposefully in a 'context of situation'. In his early work (summarised in Dixon, 1965, pp 91-97), this complex relationship is couched in the primacy of form over context. In later work (best interpreted in the early work of Kress (1994)), the relationship between form and context is explored in a more balanced way via the theory of systemic functional linguistics. A second major contribution by Halliday and his school, then, was to explore the relationships between the forms of language (e.g. lexical and syntactic elements), and the functions of language in particular contexts. The

tradition of relating text to context (Fairclough, 1992; Halliday and Hasan, 1985; Hodge and Kress, 1993) sees grammatical knowledge as serving the development of critical understanding as to how texts do their socialising work.

It is fully acknowledged in the present review that sentence-level grammar is contingent upon the notion of levels of text grammar ('above the level of the sentence') and of word grammar ('below the level of the sentence'). Nevertheless, our aim was to focus in the in-depth review on sentence-level operations in teaching about writing and in learning to write.

Key definitions

Grammar refers, as far as the present project is concerned, to written sentence and text grammars. It includes the study of syntax (rules governing word order), clause and phrase structure, and the classification of parts of speech (e.g. noun, verb, etc.), and issues regarding the cohesion and coherence of whole texts. It can be both descriptive, in that it describes the existing patterns of sentences and texts; and, in sentence terms, also generative or transformative, in that rules can be defined which can generate grammatically acceptable sentences (the transformation being from basic deep structural rules, through to actual sentences). Studies of words or sub-components of words are not part of the study of grammar *per se*. Similarly, studies in language awareness are not, strictly speaking, part of the present review, although the larger category of language awareness may come into play in considerations of grammar.

By *written composition*, we mean extended pieces of writing (in handwriting, in type or via word-processing) in a variety of genres or text-types.

In focusing on *accuracy*, we mean to place emphasis on appropriateness of grammatical form for particular purposes. We are not concerned with spelling accuracy, neither with legibility, neatness of handwriting or vocabulary (except where it bears upon sentence grammar). The emphasis on *quality* is there to distinguish our study from an interest in *quantity*.

By *English-speaking countries* we mean countries where English is spoken as a first language by a significant segment of the population[1]. We include Australia, Canada, Gibraltar, Ireland, Jamaica and other countries in the Caribbean, New Zealand, South Africa, the UK and the US; we will exclude Bangladesh, China, India, Malaysia, Pakistan and Singapore.

1.3 Policy and practice background

The teaching of grammar: the policy, practice and research contexts

Since the publication of the Kingman Report (Department of Education and Science (DES), 1988), there has been a conviction amongst curriculum writers

---

[1] We were conscious that some detailed awareness needed to be demonstrated on schools' ethnic composition when studies were data-extracted. EPPI-Reviewer enabled us to record the ethnic composition of classes and this is a factor we took into account in our narrative synthesis of results. We were inevitably constrained by whether research studies report on the ethnic composition of the classes they investigate.

and policy-makers in England that grammar teaching to young learners of English is a good thing; that it will improve their written English and their ability to talk about language; that talking about language is helpful in understanding language and, in turn, in improving its use; and that such reflection and discussion *about* language should start earlier than had previously been thought possible or desirable.

It should be said at the start that, in Perera's view, such a conviction flies in the face of research evidence. Perera (1984, p 12) notes:

> Since the beginning of the [20th] century, a body of research has accumulated that indicates that grammatical construction, unrelated to pupils' other language work, does not lead to an improvement in the quality of their own writing or in their level of comprehension. Furthermore, the majority of children under about fourteen seem to become confused by grammatical labels and descriptions. It is obviously harmful for children to be made to feel that they 'can't do English' because they cannot label, say, an auxiliary verb, when they are perfectly capable of using a wide range of auxiliary verbs accurately and appropriately. There is a brief summary of this research evidence in Wilkinson (1971, pp. 32-35).

Wilkinson notes that, although grammar is a useful descriptive and analytical tool, 'other claims made for it are nearly all without foundation' (ibid, p 32). Studies in the 20th century have suggested that the learning of formal, traditional (i.e. not transformative) grammar has no beneficial effect on children's written work (Rice, 1903); that training in formal grammar does not improve pupils' composition (Asker, 1923; Macaulay, 1947; Robinson, 1960); that ability in grammar is more related to ability in some other subjects than in English composition (Boraas, 1917; Segal and Barr, 1926); that a knowledge of grammar is of no general help in correcting faulty usage (Benfer, 1935; Catherwood, 1932); that grammar is often taught to children who have not the maturity or 'intelligence' to understand it (Macaulay, 1947; Symonds, 1931); and that teaching grammar may actually hinder the development of children's English (Macaulay, 1947).

Policy and practice in the 1970s and 1980s in England have followed a line characterised by the Bullock Report (DES, 1975), specifically that it was *teachers* who needed to know about grammatical construction so that they could understand pupils' writing problems and intervene accordingly and appropriately:

> We are not suggesting that the answer to improved standards is to be found in…more grammar exercises, more formal speech training, more comprehension extracts. We believe that language competence grows incrementally, through an interaction of writing, talk, reading and experience, the body of resulting work forming an organic whole. But this does not mean it can be taken for granted, that the teacher does not exercise a conscious influence on the nature and quality of its growth (DES, 1975, pp 7-8).

In New Zealand, emphasis has been on knowledge about language and exploring language rather than on grammar teaching *per se* (Ministry of Education, 1996). There is scepticism about the value of grammar teaching for the improvement of writing ability:

The primary purpose of this investigation was to determine the direct effects of a study of transformational-generative grammar on the language growth of secondary school pupils. The results presented show that the effects of the three years of such grammar study are negligible. Those pupils who studied no formal grammar for three years demonstrated competence in writing and related language skills equal to that shown by the pupils who studied transformational or traditional grammar. Furthermore, their attitude to English as a subject of study was more positive (Elley *et al.*, 1979, p 98).

In these respects, the English and New Zealand positions are similar: they have seen a diffusion of emphasis on grammar teaching and a resultant reorientation around language awareness. *Exploring Language: A Handbook for Teachers* states:

Knowledge of the workings of language is also essential for teachers to be able to examine and assess their students' language use in a systematic and productive way. Behind messy handwriting and creative spelling, there could well be signs of interesting language development and attempts at new complexities and variation that could pass unnoticed by those who do not have a knowledge of understanding to recognize them. How can a teacher appreciate a student's new developments with passive verbs or modal auxiliaries if these concepts themselves are not known or recognized? (Ministry of Education, 1996, p 3).

More recently, in England and Wales, the National Literacy Strategy (which operated for 7 to 11 year-olds from 1997 before being extended to 11 to 14 year-olds in 2002), has issued a book and video entitled *Grammar for Writing* (Department for Education and Employment (DfEE), 2000), aimed particularly at the teaching of 7 to 11 year-olds. The basic principle behind this relatively recent initiative is that 'all pupils have extensive grammatical knowledge' (DfEE, 2000, p 7), and that teaching that focuses on grammar helps to make this knowledge explicit. Such explicitness, so the book and video argue, helps to improve young people's writing through providing them with an increase in 'the range of choices open to them when they write' (ibid). Throughout, there is a distinction between spoken grammars and written grammars, and a clear objective to support the development of a command in sentence construction. In pedagogic terms, the emphasis of the book is on teaching at the point of composition rather than correcting after the event. While eschewing a return to the descriptive and prescriptive grammar teaching of the 1950s and 1960s, this approach does focus clearly on the improvement of sentence structure and uses extensive 'knowledge about language' and increased language awareness as a means to help pupils to write better English. It consists of a detailed programme for using sentence grammar to improve sentence construction, via explicit teaching. As such, it represents a middle ground between traditional grammar teaching on the one hand, and language awareness arising from the use of language in speech and writing on the other.

It is interesting to note that in the evaluation of the pilot of 'The Key Stage 3 Strategy' (Office for Standards in Education (Ofsted), 2002), an extension of the National Literacy Strategy from ages 7-11 to the 11-14 age group in England and Wales, the inspectors observed that in terms of attainment, improvements were clearest in spelling and stylistic conventions, and weakest 'in sentence structure, punctuation and paragraphing' (Ofsted, 2002, p 13). Such a finding would confirm

our own in this report, that there is no evidence that the teaching of grammar (syntax) nor increased language awareness have an effect on young people's syntactic maturity. And yet policy-makers continue to believe – based, we feel, on poor evidence – that teaching grammar (syntax) is beneficial. For example, enshrined in the National Curriculum (England and Wales) for English for students aged 11-16 is the following:

> Pupils should be taught the principles of sentence grammar…and use this knowledge in their writing. They should be taught:
> a. word classes or parts of speech and their grammatical functions
> b. the structure of phrases and clauses and how they can be combined

and….

> e. the use of appropriate grammatical terminology to reflect on the meaning and clarity of individual sentences (for example nouns, verbs, adjectives, prepositions, conjunctions, articles)
> (DfEE/Qualifications and Curriculum Authority (QCA) 1999, p 38).

Publications like the QCA's *Not Whether but How: Teaching Grammar in English at Key Stages 3 and 4* (QCA, 1999) assume the teaching of grammar is beneficial and do not concern themselves with the *why* or *what* questions.

Whose conventions?

The National Curriculum for England and Wales, when it was first established in the late 1980s and early 1990s, indicated that children should be able to talk about 'grammatical differences between Standard English and a non-standard variety'. Specifically, 'Standard English' refers to a broad set of conventions observed in the UK about the use of written English. Such a conception is not on the whole affected by accent. You can speak standard spoken English with a Scottish accent and written standard English is even less culturally specific. However, it has to be acknowledged that written American English has a different grammar from written British English.

Even with a broadly accepted set of conventions, there is room for disagreement and variation. Opinions about the nature of grammar, grammatical 'correctness' and the teaching (or not) of grammar make this a contentious field.

Hudson, in his book, *Teaching Grammar* (1992), suggests that 'until you know what is on the menu you can't choose from it' (p xi). In arguing the case for increased awareness of language construction amongst teachers, he is saying something similar to the Bullock Report's position that it is useful for teachers to know about grammatical construction so that they can help pupils appropriately; or Perera's (1984) similar conclusion. It may be that there is a degree of consensus among researchers and policy-makers from the 1970s to the 1990s: specifically, that, at the very least, teachers of English should know about grammar so that they can advise their pupils according to their particular needs. Perhaps a key distinction to be made at this point – one that might have a bearing on the systematic review undertaken – is how much teachers need to know about grammar in order to teach writing, and how much pupils need to know in order to write well.

Kress (1994) provides another, more radical perspective, on grammar and grammar teaching. He starts from the premise that a grammar 'is adequate if that grammar allows a speaker to express the range of meanings which that speaker needs to express in such a way as to be understood in a regular and predictable manner by a fellow user of that grammar' (Kress, 1994, p 160). In other words, a grammar is an adequate set of conventions for a particular social group or in a particular social situation; it is not a Chomskian 'universal grammar'. Thus a child's grammar may differ from an adult's and 'the whole idea of correcting a child's grammar assumes that the child's grammar is inadequate to the expression of the child's meanings' (op. cit., p 163).

The developers of *Exploring Language* (Ministry of Education, 1996) asserted that 'students and teachers need to be able to use a nationally agreed metalanguage of concepts and terminology to describe and discuss language' (p 7). In describing the process they went through to decide on this nationally agreed metalanguage, they write, 'rather than subscribing to one particular school of thought or approach to describing language, this book uses the descriptions and terminology that *will be most useful to teachers in the work with students*' (our italics). They describe this approach as eclectic. It could be argued that the writers of this book favoured Quirk *et al.* (1985) – a descriptive approach to grammar – over systemic functional grammar as the basis for their taxonomy, and therefore that they opted for a bottom-up grammar: one that does not deal with such aspects as cohesion or coherence. There is clearly a metalanguage set out in *Grammar for Writing* (DfEE, 2000), mentioned in the previous section.

Our own position in the current review is to be open to both the bottom-up approach and to the top-down approach in the systematic map of the research in the field, and then to focus on sentence grammar for the in-depth reviews. In the former case, the constructions and choices made are *informed* by semantic, textual and contextual factors. In the latter case, there is an emphasis on parts of speech and combining rules without much consideration of why certain combinations are acceptable and others not.

Research might well be needed that compares the effect/impact on students' writing skills of teachers' grammatical knowledge. It would appear that the assumption that teachers need this grammatical knowledge is more widely held than the assumption that students need to have it (to write well).

Grammar and the National Curriculum

The Kingman Report (DES, 1988), mentioned earlier, was a key document in the formulation of policy on grammar teaching and language awareness in England and Wales. Its general recommendations were to increase language awareness among pupils by increasing it among teachers at both primary and secondary levels in schooling. Although one of its recommendations – that 'by the end of the [20th] century a prerequisite for entry to the teaching profession as an English specialist should normally be a first degree which incorporates the study of both contemporary and historical linguistic form and use' (DES, 1988, p 70) – has not been met, the advent of English Language courses at Advanced Level and the development of the National Literacy Strategy are indications of an increased emphasis on language study.

The study of grammar – the forms of the language at sentence and discourse levels – is but a part of the model proposed by Kingman, which also includes three other dimensions: communication and comprehension, acquisition and development, and historical and geographical variation (ibid, pp 17ff).

The latest version of the National Curriculum for England suggests that 'pupils should be taught some of the grammatical features of written standard English' as early as Key Stage 1 (ages 5 to 7) (DfEE/QCA 1999, p 21). By Key Stage 2 (ages 7 to11), as far as reading is concerned and under the heading of 'Language structure and variation':

> To read texts with greater accuracy and understanding, pupils should be taught to identify and comment on features of English at word, sentence and text-level, using appropriate terminology (op. cit., p 26).

One example is the use of varying sentence length and structure. In writing, at this stage,

> some of the differences between standard and non-standard English usage, including subject-verb agreements and use of prepositions (op. cit., p 29)

should be taught. More detail is forthcoming on language structure, where pupils should be taught:

- word classes and the grammatical functions of words, including nouns, adjectives, adverbs, pronouns, prepositions, conjunctions, articles
- the features of different types of sentence, including statements, questions and commands, and how to use them
- the grammar of complex sentences, including clauses, phrases and connectives (ibid)

The refinement of these details at Key Stages 3 and 4 (11 to 16) simply requires that pupils should be taught 'the principles of sentence grammar…and use this knowledge in their writing'. Such teaching should include 'word classes or parts of speech and their grammatical functions' and 'the structure of phrases and clauses and how they can be combined' (op. cit., p 38). This is rather a restricted approach.

It is interesting to note that the major push on grammar teaching comes at Key Stage 2 (7 to 11). Wyse (2001) argues that the 'Grammar for Writing' initiative is insufficiently supported by empirical evidence on the teaching of grammar 'and that changes will need to be made to English curriculum policy and pedagogy if children's writing is to further improve' (op. cit., p 411). The debate continues.

By way of contrast, the 'Exploring Language' strand of *English in the New Zealand Curriculum* (Ministry of Education, 1994) suggests that pupils at levels 3 and 4 (approximately years 5-8) 'identify, discuss, and use the conventions, structures, and language features of different texts, and discuss how they relate to the topic'. At levels 5 and 6 (years 9-12), there is a greater rhetorical focus, with students expected to be 'using appropriate terminology [to] describe, discuss, analyse, and apply the distinctive conventions, structures, and language

features of a range of texts and explain how they suit the topic and purpose' (op. cit., p 36).

Policy context in the US

We have noticed that research in the US has taken rather a different line than that in the UK. This may have had something to do with an enduring interest in the US over the most effective way to teach reading rather than writing. For many years, from the early 1960s to well into the 1990s, views became polarised between advocates of a 'bottom-up' approach, who saw reading as developing incrementally from the smallest units of letters and sounds through a series of levels to the larger structures of sentences, and those who saw the acquisition of reading as a mix of skills based on the 'whole language' concepts. The latter school emphasised the contribution of contextual cues, prediction and text discourse features in the process of learning to read.

US research interest in improving writing was less by comparison. What research there was tended to reflect, from the late 1950s onwards, the rise of generative grammar as a theory of language, and used techniques (such as sentence-combining exercises) to teach and test children's acquisition of transformations.

It is also worth noting the different policy-making context that obtains in the US. The federal government has an agenda-setting role within education, and sets goals and broad aims. States have more autonomy of practice than, for instance, regions or local education authorities (LEAs) currently possess within the UK, where central government has a powerful role in setting down how teachers should teach. It is therefore feasible that states may vary widely in the significance they attach to writing quality in their state-wide education policies. The emphasis on reading, mentioned above, had explicit steers from federal aims.

The Grammar Papers

A helpful but little-publicised document, *The Grammar Papers* (QCA, 1998), provides a critical digest of the then available research into the value of teaching grammar. Aimed at teachers, it urges caution in reading too much into claims made about teaching formal grammar. It reveals that the evidence for and against explicit teaching of grammar is less reliable than it is often taken to be. It questions the assumption that writing benefits from increased knowledge of grammatical terms, pointing out that much of the pre-1960s research quoted as showing formal grammar teaching to be ineffective is judged against the expectation that it *should* be effective.

The Grammar Papers raises a number of questions for which it finds no conclusive evidence. It points out the lack of evidence about whether teaching grammar has any impact upon reading, speaking and listening as well as writing. It points out the absence of reliable evidence on the efficacy of different approaches to teaching formal grammar, on how other aspects of the curriculum are affected, on whether children's attitudes are relevant, and on the relative performances of different groups of pupils.

The document offers the view that there is no evidence to show that discrete and de-contextualised teaching of parts of speech and parsing will transfer into writing competence, although it accepts that it may well improve pupils' performances on tests of those technical abilities. It argues that it may be time to move the focus away from whether teaching grammar improves writing, towards a different reason for teaching grammar, such as that as a strand in the teaching and learning of language it counts as another tool for developing literacy. Yet this very suggestion would seem to reveal a subtext that grammar teaching is advantageous and beneficial in the curriculum.

1.4 Research background: previous systematic reviews and seminal works in the field

The first major study of the use of formal grammar in the teaching of writing was that by Macauley (1947). However, Macauley focused on the question of at what stage formal grammar should be taught, rather than whether it was appropriate and effective for it to be taught. He came to the conclusion, after a number of tests on the effectiveness of grammar teaching, that neither upper primary (i.e. 11-12 year-old) pupils, nor junior secondary (i.e. 13-14 year-old) pupils, could be depended on to recognise simple examples of nouns, verbs, pronouns, adjectives or adverbs after several years of having been taught it in English lessons (the latter group, for six years). Only upper secondary (i.e. 15-17 year-old) pupils, and those in the top boys' and girls' classes in each year, were able to reach the 50% pass standard set in Macauley's tests. His overall conclusions are that scores rise with age and schooling but that for most pupils, age and schooling are not in themselves enough for a mastery of even the most simple rules in English formal grammar; and that 'those who pass our standard are few in number and are in the best of the [upper] secondary classes' (Macauley, 1947, p 162). The implications Macauley draws out for the stages of schooling are clear: there is no point in trying to teach formal grammar in the primary years or even in the lower secondary years; it is a practice and field best reserved (if at all), for brighter pupils in the last years of secondary schooling. The study does not look at the effect of such teaching on writing accuracy or quality, but it does point out the difficulties of the first part of our research question: the teaching (and by implication, the learning) of formal grammar.

As Braddock *et al.* (1963) note, in a review of the state of knowledge about composition for the National Council of Teachers of English (US), the merit of formal grammar as an instructional aid is 'one of the most heavily investigated problems in the teaching of writing' (op. cit., p 37). They summarise the field by stating that 'study after study based on objective testing rather than actual writing confirms that instruction in formal grammar has little or no effect on the quality of student composition' (op. cit., p 37) and that 'direct methods', rather than methods based upon a knowledge of so-called related grammatical elements, are more likely to be effective.

A particularly significant study undertaken in the UK was that by Harris (1962), which compared the effect of instruction in formal grammar and functional grammar over a period of two years on the writing of 228 London pupils aged 12 to 14. This study has been seen as significant because of its longitudinal dimension and its comparison of formal grammar teaching on the one hand, and 'functional or 'direct' (i.e. no formal grammar teaching), on the other.

Harris writes in the abstract to the thesis:

> In this work, the value of the traditional English grammar lesson in helping children to write correctly was tested. The grammar lesson was found to be certainly not superior, and in most instances was inferior, to direct practice in writing skills. The progress of five forms having no grammar lesson was measured on eleven counts against that of five similar forms following the same English course but taking one lesson a week of English grammar. At the end of two academic years, of the fifty-five resultant scores, twenty-five proved highly reliable (op. cit.)

Eleven measures were used in judging essays written at the beginning and end of the experimental period: the average length of correct simple sentences (not reliable); instances of omission of the full-stop (fairly reliable); the number of words per common error (very reliable); the variety of correct sentence patterns used (very reliable); the number of correct non-simple sentences minus correct simple sentences (fairly reliable); the total number of subordinate clauses (very reliable); the total number of words (not reliable); the number of correct complex sentences minus the number of incorrect (very reliable); the number of correct simple sentences with two modifying phrases (fairly reliable); the number of total correct sentences minus incorrect (fairly reliable); and the number of adjectival phrases and clauses (fairly reliable). There were thus five very reliable measures, four fairly reliable ones, and two were not reliable.

Detailed results show that in ten out of the 25 very reliable scores, significant gains were made by the non-grammar classes (n=109), with no significant gains being made by the classes studying grammar (n=119). Specifically, 'mechanical, conventional correctness – as in the number of words per common error; maturity of style – as in the variety of sentence patterns used; the control of complex relationships – as in the number of correct complex sentences; as well as general overall correctness, seen in the total number of correct sentences, were all improved significantly in groups practising direct writing skills as compared with the groups studying formal grammar' (op. cit., p 203).

Harris is aware that the results must be treated with caution because the experimental and control groups were not strictly comparable. But he claims that there was no critical need to equate exactly the groups in each school; that the general attainment and that in English 'were roughly of the same standard' (op. cit., p 206); and that the content and order of the grammar and non-grammar syllabi were not significant 'since formal grammar itself has a vague and fluctuating meaning in present usage' (ibid).

At the time the thesis was written – and we can safely assume, for the decade or so prior to its writing – about one-fifth of English class time was devoted to the teaching of 'formal' grammar. This figure is reflected in the amount of space given to grammar instruction and exercises in textbooks at the time. Harris questions, in the light of his findings, whether such time is worthwhile, particularly as his results echo those of Macauley in that 'no real likelihood exists of successfully teaching formal English grammar to any but bright children' (Harris, 1962, p 196).

Harris therefore argues for a 'grammar of situation': that is, the study and practice of language in action rather than of the artificially narrow formal grammars.

What are the limitations of Harris' study? First, although the empirical data-gathering part of the study takes place over two years, Harris admits himself that this is the 'source…of much of the organisational fallibility' (op. cit., p 111). Second, there were only two forms running in pairs in each school, and thus the sample is relatively small. Third, it was not possible to have complete control over the experimental situation over a two-year period: 'A number of variables had to be accepted without adequate control, in the hope that the difference between the work done by the experimental groups would be sufficiently large and clear to counter-balance in the results uncertainty due to uncontrolled variables or to lack of random or representative sampling' (op. cit., p 112). Because the five schools used in the study consisted of two grammar schools, two technical/comprehensive and one secondary modern, the schools 'necessarily decided the groups of children who could be used, and in this there was no possibility of selecting two ideally equated groups, either in intelligence, background or attainment' (op. cit., p 113). In other words, although every effort was made to control the study (for example, in one teacher teaching both the control and experimental groups in each of the schools), there were variables that were not controlled. The results of the study, therefore, have to be taken with a degree of caution.

Braddock *et al*. (1963) point out that the Harris study 'does not necessarily prove…the ineffectiveness of instruction based on structural or generative grammar' (op. cit., p 83).

Tomlinson (1994) is the most critical of Harris' approach. He points out the fact that the study sample was neither randomised nor fully controlled, but accepts that such weaknesses were not decisive. More important for Tomlinson is the fact that there seems to be no clear distinction in the Harris study between the two types of grammar being taught: on the one hand, formal teaching of grammar (or indeed, teaching of formal grammar); and on the other, what appears to be more time devoted to composition but with coaching in error avoidance – what might be described as a linguistically informed process of teaching composition. The fact that the same teacher taught both experimental and 'control' classes in a single school suggests, to Tomlinson, that the 'non-grammar' class probably was in receipt of indirect grammar teaching rather than no grammar teaching. Tomlinson argues that the over-simplification of Harris' results and conclusions led to an uncritical acceptance that grammar teaching (i.e. formal, 'arid', 'parts of speech' grammar) was unproductive, and thus to policy and practice decisions that were based on a simplistic distillation of research that was itself flawed in two important respects.

Wyse (2001) defends Harris against Tomlinson's criticisms that his distinction between 'grammar' and 'non-grammar' approaches was really a distinction between a formal grammar approach and an informal grammar approach; we agree with Wyse that such a point does not invalidate Harris' findings. But we do have to accept that the Harris study was not entirely reliable.

What is interesting is how policy and practice tend to over-simplify the results of research according to the *zeitgeist* or the biases of the period. Such a phenomenon suggests that there needs to be better summarised reporting of research, with implications for policy and practice drawn out to help define exactly what these implications might be.

Two previous systematic reviews have been published in the wider field. A systematic review of findings from experimental and quasi-experimental

investigations into the effectiveness of second language instruction was published by Norris and Ortega (2000). Our review is focusing on teaching English as a first language and so the Norris and Ortega review will not be discussed further here.

Hillocks (1984, 1986) published a meta-analysis of experimental studies designed to improve the teaching of written composition. He analysed the experimental research between 1960 and 1982 on all interventions to improve written composition through a series of meta-analyses. Two of these were meta-analyses of trials of the effect of teaching grammar and sentence-combining. Hillocks concluded that grammar instruction led to a statistically significant decline in student writing ability and that this was the only instructional method of those examined not to produce gains in writing ability. Five experimental/control treatments focused on grammar in one treatment but not in the other. When compared with courses designed to teach writing tasks directly, the grammar group performed consistently worse on the essay writing exercise. The mean effect size (a given treatment gain or loss expressed in standard score units) for grammar instruction was –0.29 (CI –0.40 to –0.17). Hillocks concluded that 'every other focus of instruction examined in this review is stronger' (1984, p 160). Five studies were included in the meta-analysis that focused on sentence-combining as a method of instruction. The mean effect size for sentence-combining was 0.35 (CI 0.19 to 0.51) (statistically significant positive effect). Hillocks concluded that his research showed 'sentence-combining, on the average, to be more than twice as effective as free writing as a means of enhancing the quality of student writing' (op. cit., p 161). However, Hillocks was comparing the pooled effect sizes calculated in the meta-analyses for various interventions versus control groups, rather than pooled effect sizes for grammar interventions compared directly with other interventions.

This present systematic review is, therefore, required because the only other systematic review in the field is now twenty years out of date, and because that review did not focus exclusively on investigating the effectiveness of grammar teaching on the quality of children's and young people's (aged between 5 and 16) writing, but rather, included other populations, in particular 'college students'.

1.5 Authors, funders, and other users of the review

The authors of the present review are stated at the beginning of the report. They include researchers and a doctoral student from the Department of Educational Studies at the University of York. Two of the researchers are former Heads of English in secondary schools in the UK; one is an applied linguist. Additionally, there are researchers from Durham (UK), and Waikato (New Zealand) universities, one of whom held senior posts in primary education and the other in secondary education. Furthermore, there is an experienced Information Officer on the review team.

The review has been funded by the Department for Education and Skills (DfES) via the EPPI-Centre at the Institute of Education, University of London, and by the Department of Educational Studies at the University of York.

The Department of Educational Studies is developing its links with schools interested in research in 2003/2004 (see Department Plan, available from Alison

Robinson). Such links will enable more teachers than those on the Advisory Group to comment on, contribute to and disseminate the work of the English Review Group. In addition, following a meeting with the Teacher Training Agency (TTA) and Post-graduate Certificate in Education (PGCE) students in June 2003, PGCE tutors and students will be involved in a pilot project to write summaries of the present research review (and previous reviews) and to prepare sample lessons arising from the research findings. In addition, it is hoped that a pupil from a secondary school in York will work on a pupil summary of the final review. The dissemination strategy of the English Review Group was discussed at the steering group meetings in September 2003 and February 2004.

User perspectives on the review(s) will be written by teachers, teacher educators, students, governors and policy-makers. Representatives from each of these constituencies (except students) have contributed to the direction and design of the review through the English Review Group's advisory steering committee.

1.6 Research questions

The initial review research question is:
What is the effect of grammar teaching in English on 5 to 16 year olds' accuracy and quality in written composition?

The conceptual framework for the review was based on the premise that sentence-level grammar is contingent upon the notion of levels of text grammar ('above the level of the sentence') and of word grammar ('below the level of the sentence'). Nevertheless, our aim was to focus in the in-depth review on sentence-level operations in teaching about writing and in learning to write.

The review includes descriptive mapping, which identifies and broadly characterises the studies, prior to the in-depth review on the effect of teaching of syntax on the quality and accuracy of 5 to 16 year-olds' written composition.

2. METHODS USED IN THE REVIEW

Systematic review methods were used throughout the review, using the EPPI-Centre guidelines and tools for conducting a systematic review (EPPI-Centre, 2002a, 2002b and 2002c).

2.1 User involvement

2.1.1 Approach and rationale

As outlined in detail in section 1.5, the English Review Group has involved teachers, school governors, teacher trainers and advisory teachers in its Advisory Group, which has commented on and supported the review at each stage. In addition, it is hoped that the results of the review will be disseminated more widely, through the user summaries, press releases and a journal article.

2.1.2 Methods used

User perspectives on the review(s) will be written by teachers, teacher educators, students, governors and policy-makers. Representatives from each of these constituencies (except students) have contributed to the direction and design of the review through the English Review Group's advisory steering committee. Following a meeting with the TTA and PGCE students in June 2003, PGCE tutors and students will be involved in a pilot project to write summaries of the present research review (and previous reviews) and to prepare sample lessons arising from the research findings. In addition, it is hoped that a pupil from a secondary school in York will work on a pupil summary of the final review.

2.2 Identifying and describing studies

2.2.1 Defining relevant studies: inclusion and exclusion criteria

For the mapping stage, we looked at primary research and reviews in the field published between 1900 and the present. We limited the review to the teaching of English grammar in schools where English is being taught as a first language (not foreign or second or additional language) in English-speaking countries. We included research with pupils aged between 5 and 16 and in full-time education. We focused on the effects of teaching of any kind of grammar on writing with either process or quantitative outcomes relating to writing quality and accuracy, and excluded studies that focused on any effects on reading, or on language acquisition, or oracy. We included both published and unpublished (but in the public domain) research, but we excluded unpublished PhD and masters theses. The reason for this was pragmatic: we simply did not have the resources to locate and send for the large numbers of PhD theses in the field that have been undertaken in the US. However, by including both published and unpublished

(but publicly available) research, we have limited the potential effects of publication bias.

As the focus of the study is on the *effects* of grammar teaching, papers using methods to identify any such effects were required. This implies the following study types, classified according to the EPPI-Centre taxonomy of study type contained in its core keywording strategy (EPPI-Centre, 2002a):

B: Exploration of relationships
C: Evaluation (naturally-occurring or researcher-manipulated)
E: Review (systematic or other review) containing at least one study exploring
   relationships or one evaluation

The full inclusion/exclusion criteria for the review are contained in Appendix 2.1.

2.2.2 Identification of potential studies: search strategy

Reports were identified from the following sources:

* searching of electronic bibliographic databases: ERIC, PsycINFO and SSCI
* citations from reference lists of all included systematic and non-systematic reviews
* personal contacts

Keywords for searching
Keywords for searching included the following:

* composition, writing, written composition
* grammar, syntax, text grammar, sentence grammar
* metalinguistics
* knowledge about language (KAL)

Appendix 2.2 contains the full search strategy for ERIC, PsycINFO and SSCI. Searches of these sources were limited so as to identify studies conducted in the time period 1900 to the present. The Review Group set up a database system, using EndNote bibliographic software, for keeping track of, and coding studies, found during the review. Titles and abstracts were imported and entered manually into the database.

2.2.3 Screening studies: applying inclusion and exclusion criteria

We applied the inclusion and exclusion criteria in three stages: successively to (i) titles and abstracts (by CJT and RJA with QA screening supplied by the EPPI-Centre), and (ii) full reports (firstly by CJT and RJA, and secondly, by all members of the review team with QA screening supplied by the EPPI-Centre). We obtained full reports for those studies that appeared to meet the criteria, or where we had insufficient information to be sure. We re-applied the inclusion and exclusion criteria to the full reports and excluded those that did not meet these initial criteria.

2.2.4 Characterising included studies

The studies remaining after application of the criteria were keyworded using the EPPI-Centre's core keywording strategy (EPPI-Centre, 2002a) and online database software, *EPPI-Reviewer* (EPPI-Centre, 2002b). Additional review-specific keywords which are specific to the context of the review were added to those of the EPPI-Centre, with definition of the terms in the glossary. The EPPI-Centre's core keywords and the review-specific keywords are contained in appendices 2.3 and 2.4 respectively.

All the keyworded studies were uploaded from EPPI-Reviewer to the EPPI-Centre's Research Evidence in Education Library (REEL), for others to access via the website.

2.2.5 Identifying and describing studies: quality assurance process

Screening
Due to the difficulties associated with screening this database, it was necessary to screen inclusively at the first and second stages and add an additional, third stage of screening.

Quality assurance was undertaken at each of the three stages of the screening process.

(i)   Screening of titles and abstracts
CJT screened the database created by the electronic searches of ERIC and PsycINFO (4,400 references); RJA independently double-screened a randomly generated 10% sample of this database (440 references). A representative from the EPPI-Centre (DE), independently screened a randomly generated 5% sample of the 10% sample. Cohen's Kappa measures of agreement were calculated between CJT and RJA (for 10% sample), and between CJT and RJA and DE (for 5% sample of 10% sample).

(ii)   Screening of full papers (first stage CJT and RJA)
CJT screened all of the papers sent for as a result of the first stage screening of titles and abstracts; RJA independently double-screened 50% of the full papers. The agreement between the two reviewers was analysed; any disagreements were discussed and resolved.

(iii)   Screening of full papers (second stage CJT, RJA, SB, JG, TL, GL, DZ)
All papers included at the second stage screening were independently double-screened by pairs of reviewers and an EPPI-Centre representative. At this stage, all decisions to include or exclude at third stage were discussed and resolved by at least two reviewers.

Keywording
Quality assurance of keywording began with a moderation exercise. All members of the review team (CJT, RJA, SB, TL, GL and DZ) and a representative of the EPPI-Centre (JG), independently keyworded two papers using the EPPI-Centre keywording strategy (EPPI-Centre, 2002a), and a draft version of keywords designed specifically for this review. The results of the moderation exercise were analysed (CJT) and used as a basis for discussion about generic and review-

specific keywording procedures. In addition, the review-specific keywords were re-drafted and additional information was provided for the glossary.

Subsequently, keywording of all the studies included in the systematic map were conducted by pairs of Review Group members, working first independently and then comparing their decisions before coming to a consensus. Members of the EPPI-Centre also helped in applying criteria and keywording a sample of studies.

2.3 In-depth review

2.3.1 Moving from broad characterisation (mapping) to in-depth review

Our protocol for the review of the effectiveness of grammar teaching on writing quality and accuracy focused on two levels of language description: sentence-level and text-level. Although we have been aware throughout that all levels of language description are inter-related, and that each level (e.g. morphological, sentence-based, text-based, etc.) relates to the next one 'up' and the next one 'down', we have tried to maintain focus on these two key levels as far as written composition is concerned.

The processes of searching, screening (using our pre-established research question and inclusion/exclusion criteria) and keywording led to the inclusion of studies fulfilling our inclusion criteria in the systematic map of research in the field at both sentence-level and text-level. We then decided, at this mapping stage, to focus on two aspects of sentence-level grammar teaching for two in-depth reviews to be undertaken in 2004 and 2005: one focusing on the teaching of syntax, and the other focusing on the teaching of sentence-combining. We decided not to focus on the teaching of text-level grammars.

There are several reasons for the decision not to pursue text-level grammar teaching in the two in-depth reviews.

First, the mapping of the field resulted in a widely varied set of studies. At one end of the spectrum were studies that focused on sentence-combining (e.g. McAfee, 1981; Willig, 1985), and at the other end, studies that focused on 'expository text structure use' (e.g. Gordon, 1990). In between, studies looked variously at instruction for deaf students in syntactic cohesion (McNeill, 1994), metacognitive learning strategies for improving paragraph writing for students with mild learning disabilities (Welch, 1992), and oral drills and writing improvement (Miller and Ney, 1967). Some of these studies (e.g. McAfee, 1981; Miller and Ney, 1967; Satterfield and Powers, 1996; Willing, 1985) appear in the map as focusing on both sentence-level grammar and text-level grammar.

Second, we are not confident (with hindsight) that our protocol, searching strategy and keywording approach has revealed the full extent of studies in the field of text grammar. Our pre-established inclusion/exclusion criteria specifically excluded studies exploring 'word' level grammar (morphology) and story and genre 'grammars'. At all stages of searching and screening, these levels of 'grammar' were not specifically searched for, and if retrieved, specifically excluded (Exclude 1). We know of work undertaken in Australia by the 'genre school' in the 1980s by Frances Christie and others that was not retrieved by the

search strategy; and of subsequent work in the 1990s by Freedman, Medway and others on genre as social action. Although there were good reasons for such exclusion (in that much of the work was not concerned with the teaching of grammar), we consider this work important in the field and necessary to an underpinning of theory at 'text' level. Therefore, it seems probable that some studies addressing the effect of 'text' level grammar teaching on the quality and accuracy of pupils' written composition may have been inadvertently omitted from our systematic map as we sought to exclude word and story/genre levels of grammar teaching, and due to the inevitable intertwining of 'text' level within story/genre level grammar. In this sense, sentence- and word-level grammars are more distinct.

Third, the field of 'text grammar' is a complex one, often ill-defined. There was much seminal work in the early 1980s by van Dijk and others that attempted to build on the work of Propp (1968), Rumelhart (1975), and Mandler and Johnson (1977) on story grammar. Story grammar work itself was an attempt to project transformational and generative sentence grammars to the text-level. Its tree diagrams and other forms of representation ended up shedding less light on story and narrative, and more on artificial intelligence programming – and the project itself ran into the ground. Building text grammars, as van Dijk tried to do, to account for variation and structuring of texts, proved an idealistic and impossible task. Our limited search found that the field is still ill-defined and that research in it is not coherent.

Lastly, there was a positive reason for pursuing two branches of the sentence grammar teaching research. One branch – the teaching of 'traditional' syntax-based grammar – seemed to sit alongside the other – the teaching of sentence-combining – without proper consideration of the relationship of the relative effectiveness of each approach. The first appears to have had precedence in the UK, the second in the US (particularly in the 1960s and 1970s). Undertaking two such in-depth reviews would enable us to take stock of the research in the two sub-areas of sentence grammar, thus providing a basis for further research in the field, as well as a foundation for future policy and practice.

The criteria for including and excluding studies for the in-depth review on the effect of teaching 'syntax' were refined after the systematic map was drawn. The inclusion criteria for in-depth review focused on selected review-specific keywords in order to identify studies that look at the effects of *sentence* grammar teaching (specifically syntax) on the quality and accuracy of pupils' writing.

Inclusion criteria
- Must be a study focusing on the teaching of syntax
- Must be study type B or C

2.3.2 Detailed description of the studies in the in-depth review

Studies identified as meeting the inclusion criteria were analysed in depth, using the EPPI-Centre's detailed data-extraction guidelines (EPPI-Centre, 2002c), together with its online software, EPPI-Reviewer (EPPI-Centre, 2002b).

2.3.3 Assessing the quality of studies and weight of evidence for the review question

Three components were identified to help in making explicit the process of apportioning different weights to the findings and conclusions of different studies. Such weights of evidence are based on the following:

(A) the soundness of studies (internal methodological coherence), based upon the study only
(B) the appropriateness of the research design and analysis used for answering the review question
(C) the relevance of the study topic focus (from the sample, measures, scenario, or other indicator of the focus of the study) to the review question
(D) an overall weight taking into account (A), (B) and (C)

To get from A, B and C, to D, discussion took place between reviewers. Where there was disagreement between reviewers, a third party was asked to mediate.

2.3.4 Synthesis of evidence

The data were then synthesised to bring together the studies which answer the review question and which meet the quality criteria relating to appropriateness and methodology. A narrative synthesis was undertaken. It was not felt to be appropriate to conduct a statistical meta-analysis.

2.3.5 In-depth review: quality assurance process

Data-extraction and assessment of the weight of evidence brought by the study to address the review question was conducted by pairs of Review Group members, working first independently, and then comparing their decisions before coming to a consensus. Members of the EPPI-Centre helped in data-extraction and quality appraisal of a sample of studies.

3. IDENTIFYING AND DESCRIBING STUDIES: RESULTS

3.1 Studies included from searching and screening

Table 3.1 gives the origin of all papers found and those subsequently included in the systematic map. Table 3.2 describes the identification of single studies or reviews that were reported in more than one paper. Figure 3.1 illustrates the process of filtering papers from searching to mapping and finally to synthesis.

Table 3.1: Origin of included papers

	Found	Included
ERIC	2,557	39
PsycINFO	1,844	1
SSCI	119	3
Citation	43	13
Contact	3	2
Total	*4,566*	*58*

Papers found on ERIC, PsycINFO and SSCI were imported and de-duplicated hierarchically into the review database. This is reflected in the higher proportion of included papers shown as retrieved from ERIC and included in the map. An unusually large number of studies were identified from handsearching the bibliographies of the included systematic and non-systematic reviews, as reflected in the proportionately high number of citations included in the map. Any potentially relevant studies identified through handsearching the reviews were sent for and then screened using the inclusion/exclusion criteria. Any studies that met our inclusion criteria were keyworded and included in the descriptive map (n=13).

It was necessary to keep the database open long enough in order to retrieve all the papers identified from handsearching. The database closed on Friday 9 January 2004. Any papers received after that date will be included in the update.

Table 3.2: Type of research and number of studies reported by included papers

Research type	Number of papers	Number of reviews or studies
Reviews	25	24
Primary research	33	31

The screening process identified 58 papers that met the inclusion criteria. Table 3.2 shows that 25 papers reported reviews and 33 reported primary research. One review is reported in two formats: as the full review published in a book and as a summary in a journal article (Hillocks, 1984, 1986). In addition, four papers (Combs, 1976, 1977; Elley, 1975, 1979) reported two studies.

The balance of the map therefore describes 24 reviews and 31 studies.

Figure 3.1: Filtering of papers from searching to map to synthesis

1. Identification of potential studies

Papers identified
where there is not immediate screening (e.g. electronic searching, where criteria for exclusion is recorded)
N = 4,736

Duplicate references excluded
N = 216

Criteria
1 N = 3,364
2 N = 245
3 N = 298
4 N = 300
5 N = 61
6 N = 0
7 N = 42

One-stage screening:
Papers identified in ways that allow immediate screening (e.g. handsearching, personal contact where criteria for exclusion is not recorded)
N = 46

Abstracts and titles screened
N = 4,520

Papers excluded
N = 4,310

2. Application of inclusion/ exclusion criteria

Potential includes
N = 256

Papers not obtained
N = 12

Criteria
1 N = 108
2 N = 5
3 N = 11
4 N = 35
5 N = 0
6 N = 0
7 N = 0

Papers excluded 2nd stage
N = 159

Systematic map (reviews)
N = 25 papers containing 24 reviews

Full document screened
N = 244

Papers excluded 3rd stage
N = 27

Criteria
1 N = 13
2 N = 3
3 N = 6
4 N = 4
5 N = 1
6 N = 0
7 N = 0

3. Characterisation

Systematic map (primary research)
N = 33 papers containing 31 studies

In map but excluded from in-depth review
N = 21

4. In-depth review

In-depth review
Studies included
N = 10

The effect of grammar teaching (syntax) in English on 5 to 16 year olds' accuracy and quality in written composition

3.2 Characteristics of the included studies (systematic map)

Figure 3.2: Publication dates of reviews and studies
(Reviews: N = 24, mutually exclusive)
(Primary studies: N = 31, mutually exclusive)

Figure 3.2 defines the publication dates[2] of the included reviews and studies. There was nothing identified before 1966, probably as a result of the electronic searching approach and the particular nature of the question we set ourselves. However, earlier studies were addressed in the background.
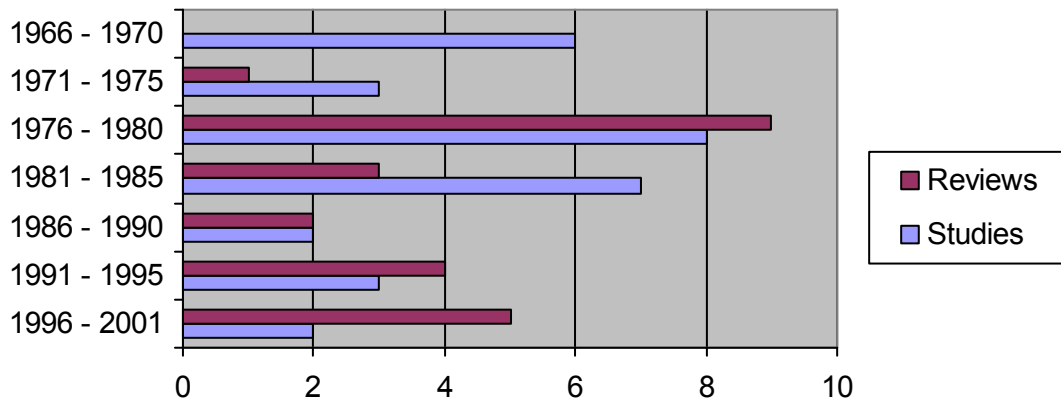
It is interesting to note that 38% (n = 9) of the included reviews were conducted in the five-year period between 1976 and 1980. This contrasts with the ten-year period between 1981 and 1990 when only 21% (n = 5) of included reviews were carried out. However, the proportion starts to rise again between 1991 and 2001, during which ten-year period a further 38% of included reviews were undertaken. Only one review included in the map was conducted prior to 1976.

Almost half (n = 15) of the primary studies included in the map were conducted in the ten-year period between 1976 and 1985. A further nine of the included studies (29%) were carried out prior to 1976, with six (19%) of these being conducted in the five-year period between 1966 and 1970. Only 23% (n = 7) of studies included in the map were conducted later than 1985.

A possible explanation for the pattern of publication observed above in Figure 3.2 is the interest in Hunt's theoretical work on T-units and the Subordinate Clause Index (S-C-I) in the 1960s (Hunt, 1966). Many of the primary studies used Hunt's S-C-I as an outcome measure.

---

[2] For the purpose of Figure 3.2 'publication date' is defined as the date that the review or study entered the public domain. As described later in this chapter, a large proportion of the included reviews and studies are in the form of research reports that are unpublished in the sense that they are available only in online databases, such as the ERIC, rather than as journal articles, books, book sections or other conventional media of publication.

All the included primary studies were study type C (i.e. evaluations); this is interesting. We searched for, and screened for, study type B (exploration of relationships) but found none.

Further characteristics of included reviews

Table 3.3: Type of review (N = 24, mutually exclusive)

Type of review	Number of reviews
Systematic	1
Non-systematic	23
Total	*24*

Almost all the reviews included in the map were non-systematic. Only one (Hillocks 1984, 1986) was a systematic review.

Table 3.4: Country of origin in which the studies were carried out (N = 24, mutually exclusive)

Country	Number of reviews
USA	15
UK	3
Canada	3
Japan*	1
New Zealand	1
Not stated	1
Total	*24*

*Undertaken by UK academic in Japanese university (Tomlinson, 1994) commenting on the scene in the UK – thus included

More than half the reviews (63%) were conducted in the US. One review in eight (n = 3) was conducted in the UK and the same number originated from Canada.

Table 3.5: Publication status (N = 24, mutually exclusive)

Status	Number of reviews
Published	14
Unpublished	10
Total	*24*

Although the majority (58%) of reviews were published, Table 3.5 shows that a high proportion (42%), were in the form of unpublished research reports.

Figure 3.3: Publication status by country of origin (N = 24, mutually exclusive)
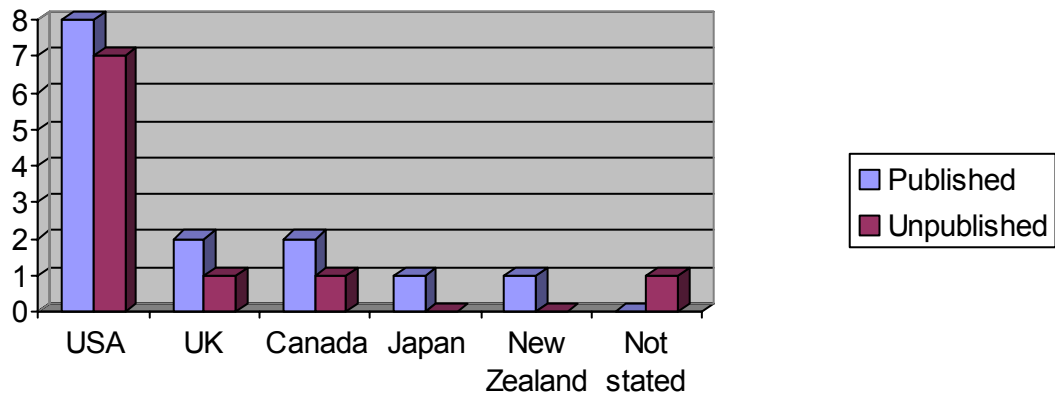
The cross-tabulation in Figure 3.3 shows that the status of reviews conducted in the US was split almost equally between published (n = 8), and unpublished (n = 7). Of those conducted in the UK and Canada, two of the three studies included for each country were published. Each of the studies conducted in Japan and New Zealand were published. The country of origin of the remaining study (unpublished) was not stated.

Table 3.6: Type of grammar teaching (N = 24, mutually exclusive)

Type of grammar	Number of reviews
Sentence-level	23
Text-level	1

In Table 3.6, we see that 23 of the 24 of the reviews included in the map reported on the teaching of sentence-level grammar. Only one review (Seidenberg, 1989), reported on text-level grammar teaching.

Table 3.7: Focus of sentence-level reviews (N = 23, not mutually exclusive)

Focus	Number of reviews
Sentence-combining	20
Syntax	16
Sentence-diagramming	1

Of the 23 reviews that reported on sentence-level grammar teaching, 20 focused (not exclusively) on sentence-combining and 16 focused on syntax (again not exclusively). Twelve reviews reported on both sentence-combining and syntax and one review reported on sentence-combining, syntax and sentence-diagramming.

Figure 3.4: Reviews of sentence-level teaching by publication status
(N = 23, not mutually exclusive)
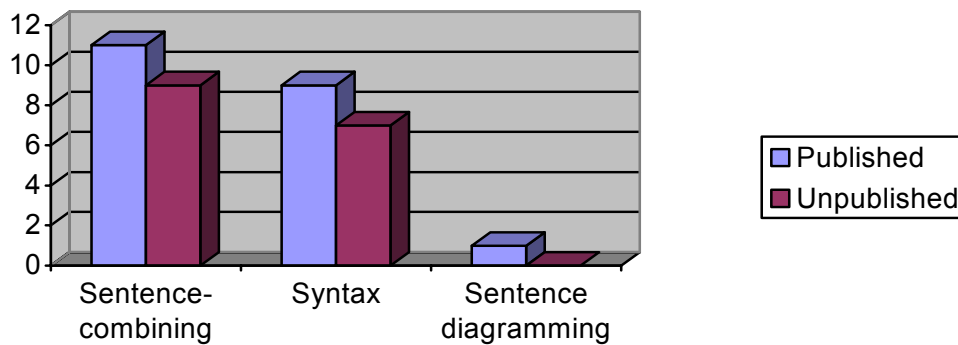
The cross-tabulation in Figure 3.4 defines the focus of the reviews on sentence-level grammar teaching by publication status. Just over half (n = 11) of the 20 reviews on sentence-combining were published. Similarly, of the 16 reviews on syntax, nine were published and seven were unpublished. The one review that reported on all three elements of sentence-level teaching was published. The in-depth review focuses on the effect of teaching syntax. Therefore the conclusions of the 16 reviews that focus on syntax are presented below.

Table 3.8: Summary of conclusions of the one systematic syntax review and 15 non-systematic syntax reviews

Author, date	Definition of 'grammar' (syntax)	Number of studies in review	Conclusion – effect of teaching grammar on writing
*Hillocks 1984 1986 *Systematic review	'The study of parts of speech, and sentences'	5 (grammar) 5 (sentence-combining)	Grammar: 'every other focus of instruction examined in this review is stronger' (1984, p 160). 'Sentence-combining': 'on the average, ..(is) ..more than twice as effective as free writing as a means of enhancing the quality of student writing' (1984, p 161).
Abrahamson, 1977	Not stated	8 evaluative abstracts, but 7 empirical studies	'…the study concludes that traditional grammar instruction does not help students improve their writing ability appreciably, that such instruction, in fact, may hinder the development of students as writers, and that sentence-combining instruction should be incorporated into both elementary and secondary language arts programs' (p 1).
Bamberg, 1981	Not stated	4	'Found grammar study does not improve writing' (pp 4-5).
Crowhurst, 1980	Not stated	8 interventions but 7 relevant studies	'First, neither T-unit length nor clause length is a good predictor of writing quality. Second, although sentence-combining studies sometimes seem to improve writing quality, the improvement is probably due to factors other than increases in T-unit and clause length' (p 2).

Author, date	Definition of 'grammar' (syntax)	Number of studies in review	Conclusion – effect of teaching grammar on writing
Elley, 1994	Systematic instruction in the analysis of sentences	5	'Formal grammar instruction appears to contribute nothing to the development of writing and reading skills' (p 1470).
Gann, 1984	Not stated	2	'…grammar instruction almost certainly does not contribute significantly to improvement in written English' (p 49).
Hudson, 2000	Not stated	13 reviews, 28 further separate studies	Hudson declares at the end that 'the idea that grammar teaching improves children's writing skills is much better supported by the available research than is commonly supposed' but his review shows that traditional grammar teaching is ineffective, on the whole; and that sentence-combining is effective.
Kolln, 1996	Not stated	4	No conclusions
Matzen *et al.* 1995	Not stated	5	No conclusions
Newkirk, 1978	Not stated	2	No conclusions
QCA, 1998	Not stated 'Grammar teaching is a complex issue'	10	'Discrete teaching of parts of speech and parsing in de-contextualised form is not a particularly effective activity…There is no evidence that knowledge acquired in this way transfers into writing competence'. 'Transformational-generative grammar…has little to offer…' 'There is evidence from studies of writing development that experience of the syntactic demands of different types of tasks is a key factor in pupils' written performance and development' (p 55).
Tomlinson, 1994	Only in discussion of one of the two papers it reviews (Harris, 1962), where it refers to 'traditional formal grammar', i.e. the 'rigidly heavy taxonomic traditional grammar, which, for example, went into four classes of adjective' (p 25)	2	'To conclude, on the basis of teaching parts of speech to 12- and 13 year-olds that grammar teaching, even the teaching of traditional grammar, had no value in secondary school, is, if not specious nonsense, certainly a non-sequitur' (p 26).
Sternglass, 1979	Not stated	2	No conclusions
Ulin and Schlerman, 1978	Not stated	4	'…these studies have shown that TG [transformational grammar] is no more effective than traditional grammar in improving composition…' (p 65).
Walsh, 1991	'…the system of word structures and word arrangements of the language' (p 3)	13	'…it does not follow that knowledge of grammar will make one a better writer' (p 7).
Wyse, 2001	It cites five definitions of grammar by Hartwell (1985)	15 reviews, 12 individual	'The findings from international research clearly indicate that the teaching of

Author, date	Definition of 'grammar' (syntax)	Number of studies in review	Conclusion – effect of teaching grammar on writing
	and refers to structural, transformational, generative, functional grammars, etc. 'This paper focuses attention on some of the empirical evidence in relation to Traditional School Grammar (TSG), transformational grammar, and sentence-combining' (p 412).	studies	grammar (using a range of models) has negligible positive effects on improving secondary pupils' writing' (p 422). 'The one area where research has indicated that there may be some specific benefit for syntactic maturity is in sentence-combining' (p 423).

The conclusions in these reviews are used to contextualise our results in the discussion section of Chapter 5.

Further characteristics of included primary studies

Table 3.9: Type of evaluation (N = 31, mutually exclusive)

Type of evaluation	Number of studies
Researcher-manipulated	30
Naturally-occurring	1
Total	*31*

Table 3.9 shows that almost all the evaluations of primary research included in the map were researcher-manipulated. Only one evaluation was found to be naturally-occurring.

Figure 3.5: Type of researcher-manipulated evaluation (N = 30, mutually exclusive)
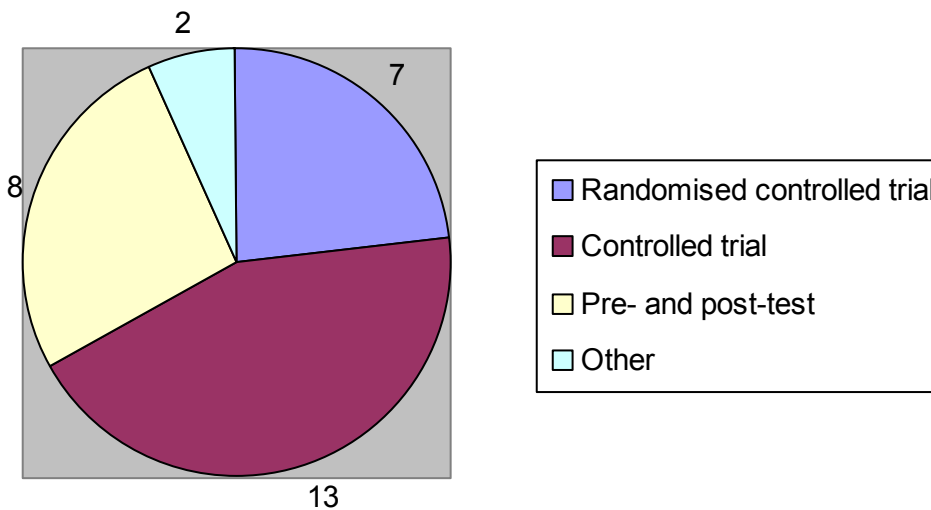
Of the 30 researcher-manipulated evaluations, seven reported randomised controlled trials, 13 reported controlled trials, eight reported pre-and post-tests, and two reported other types of evaluation.

Table 3.10: Country of origin (N = 31, mutually exclusive)

Country	Number of studies
USA	28
Canada	2
New Zealand	1
Total	*31*

Table 3.10 shows that 90% of primary studies (n = 28) included in the map originated in the US. Two studies were conducted in Canada and one in New Zealand. It is interesting to note that none of the studies were conducted in the UK.

Table 3.11: Publication status (N = 31, mutually exclusive)

Status	Number of studies
Published	17
Unpublished	14
Total	*31*

As in the case of the reviews, the majority of studies (55%) were published, but a high proportion (45%) were unpublished.

Figure 3.6: Publication status by country of origin (N = 31, mutually exclusive)
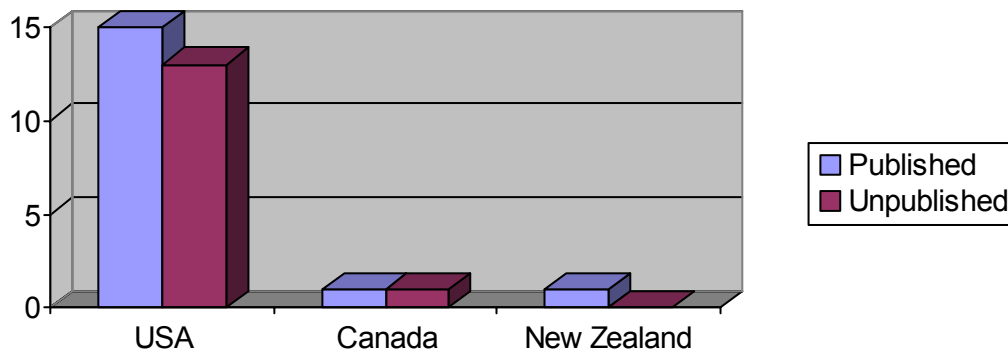
The cross-tabulation in Figure 3.6 shows that the status of primary studies originating in the US was again split almost equally between published (n = 15), and unpublished (n = 13). Of the two studies conducted in Canada, one was published and one was unpublished. The remaining New Zealand study was published.

Table 3.12: Types of learners (N = 31, not mutually exclusive)

Educational setting	Number of studies
Primary school	12
Secondary school	15
Special needs school	3
Independent school	2
Residential school	1

Age of learners	
5-10	16
11-16	21
Sex of learners (mutually exclusive)	
Mixed sex	15
Male only	2
Female only	1
Not stated	13

Table 3.12 describes the educational settings in which the studies were conducted and the age and sex of the learners involved. Two studies were conducted in more than one educational setting and six studies involved learners in both primary and secondary school age groups. Half the studies (n = 15) involved learners of mixed sex. In 13 studies, the sex of learners was not stated.

Table 3.13: Type of grammar teaching (N = 31, not mutually exclusive)

Type of grammar	Number of studies
Sentence-level	28
Text-level	9

In Table 3.13 we see that 28 studies (90%) focused on the teaching of sentence-level grammar. Nine studies focused on text-level teaching and six involved the teaching of both types of grammar.

Table 3.14: Focus of sentence-level studies (N = 28, not mutually exclusive)

Focus	Number of studies
Sentence-combining	20
Syntax	10
Punctuation	3
Sentence-diagramming	1

Of the 28 studies that reported on sentence-level grammar teaching, 20 focused on sentence-combining and ten focused on other aspects of syntax. A much smaller proportion focused on punctuation (n = 3), and only one study focused on sentence-diagramming. Three studies investigated the teaching of both sentence-combining and syntax. One study focused on sentence-combining and punctuation, one on syntax, punctuation and sentence-diagramming, and one on punctuation alone.

Figure 3.7: Studies of sentence-level teaching by publication status
(N = 28, not mutually exclusive)
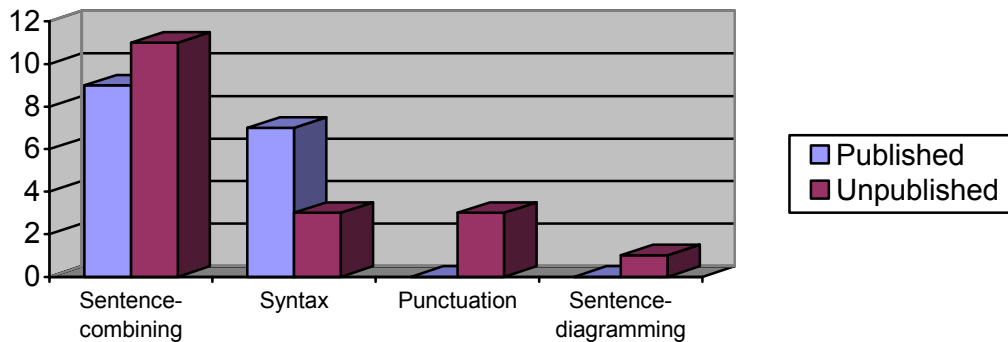
It is interesting to note from the cross-tabulation in Figure 3.7 that, of the 20 studies focusing on sentence-combining, just over half (n =11) were unpublished. The reverse was true of the studies concerned with syntax. In this group, seven were published and three unpublished. The balance of the studies on punctuation (n = 3) and sentence-diagramming (n = 1) were unpublished.

Table 3.15: Focus of text-level studies (N = 9, not mutually exclusive)

Focus	Number of studies
Paragraph composition	6
Text structure	4
Cohesion	2

Of the nine studies that reported on text-level grammar teaching, two-thirds (n = 6) focused on paragraph composition. Four studies focused on text structure and two on cohesion. Two studies investigated the teaching of both paragraph composition and text structure, and one study investigated both paragraph composition and cohesion.

Figure 3.8: Studies of text-level teaching by publication status
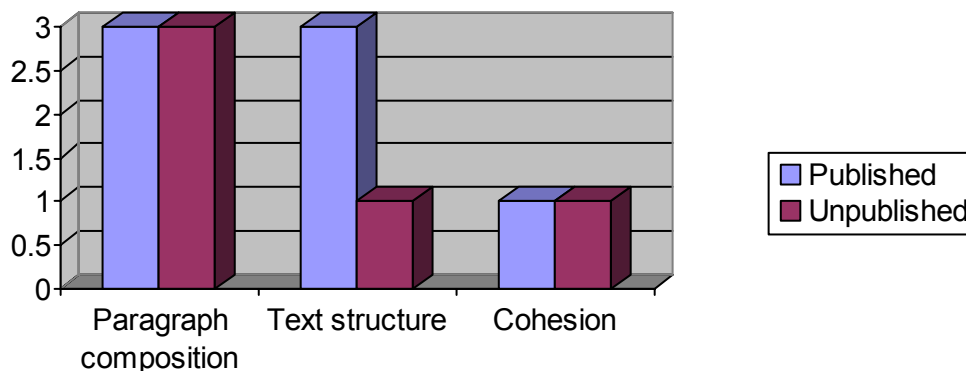(N = 9, not mutually exclusive)

Figure 3.8 shows that, of the six studies concerned with paragraph composition, half were published and half were unpublished. One of the three studies focusing

on text structure was unpublished and, of the two studies that focused on cohesion, one was published and one was unpublished.

3.3 Identifying and describing studies: quality assurance results

Quality assurance of the three stages of screening papers retrieved from the electronic searches

Screening
(i)   *Screening of titles and abstracts*
The Cohen's Kappa measure of agreement between CJT and RJA for screening of the 10% random sample of the full database was **0.81 (very good)**. The Cohen's Kappa measure of agreement between CJT and RJA for screening the 5% random sample of the 10% sample was **1 (perfect agreement)**. The Cohen's Kappa measure of agreement between DE (EPPI-Centre) and CJT (and RJA) for screening the 5% random sample was **0.64 (good)**.
CJT and RJA adopted an inclusive approach to screening because of the difficulties encountered with the database (i.e. the difficulty of distinguishing between different levels of grammar: 'word', 'sentence', 'text' or 'genre'). The lower level of agreement arose because DE adopted a slightly more rigorous approach to screening.

(ii)   *Screening of full papers (first stage CJT and RJA)*
The papers identified as being potentially relevant from the screening of the database were sent for and re-screened on the basis of the full papers by CJT. One hundred and fifty papers were excluded at this stage. Fifty percent (n = 75) of the papers excluded at this stage were re-screened by RJA. There was disagreement on two papers; this disagreement was resolved through discussion. Because the agreement between the two reviewers was high (agreement to include or exclude on 73 out of 75 papers), it was felt that the review team could have confidence in the main screener at the second stage and it was not felt necessary to double-screen the remaining 50% of papers (n = 75).

(iii)   *Screening of full papers (second stage CJT, RJA, SB, JG, TL, GL, DZ)*
Because of the difficulties of screening the database, this additional stage of screening was adopted. The difficulties of screening papers for inclusion in this review were due to problems with ascertaining the level of grammar teaching operating in the studies (word, sentence or text). Therefore initial inclusive screening at first and second stages led to possible exclusions at the third stage. At this stage, two reviewers independently read all the papers, discussed them and finally resolved all decisions with regard to inclusion in the map.

Keywording
Keywording moderation was based on two papers: Bateman and Zidonis, 1996; and Combs, 1976. Full agreement was established on all generic keywords, except number 3 (i.e. whether or not the study was linked). Good agreement was established on most of the review-specific keywords. In the light of feedback from reviewers, the review-specific keywords were re-drafted (addition of different types of grammar) and the definitions were clarified for contextualised and de-contextualised grammar teaching.

4. IN-DEPTH REVIEW: RESULTS

4.1 Selecting studies for the in-depth review

Ten studies were identified for the in-depth review. These studies were identified through the application of the review-specific keyword 'syntax' to the primary studies in the map.

4.2 Further details of studies included in the in-depth review

The ten studies selected for in-depth review were all researcher-manipulated experimental studies, of which two were randomised controlled trials (Bateman and Zidonis, 1966; Fogel and Ehri, 2000); two were controlled trials (Elley *et al.*, 1975, 1979; Stock 1980); four used pre- and post-tests (Hilfman, 1970; McNeill, 1994; Roberts and Boggase, 1992; Rousseau and Poulson, 1985); one was a curriculum evaluation (Satterfield and Powers, 1996); and one a single subject ABACA design (Stone and Serwatka, 1982).

Full details of all the included studies are contained in the summary tables in Appendix 4.1.

4.3 Synthesis of evidence

The overall weight of evidence for the studies was as follows.

Table 4.1: Weight of evidence for studies included in the in-depth review

Study	Overall weight of evidence
Bateman and Zidonis (1966)	Medium to high
Elley *et al.* (1975, 1979)	High to medium
Fogel and Ehri (2000)	High
Hilfman (1970)	Medium to low
McNeill (1994)	Medium to low
Roberts and Boggase (1992)	Low
Rousseau and Poulson (1985)	Medium to low
Satterfield and Powers (1996)	Low
Stock (1980)	Medium to low
Stone and Serwatka (1982)	Low

The overall weight of evidence for the studies included in the in-depth review was calculated by reference to the study's trustworthiness in relation to the study questions, the appropriateness of the research design and analysis, and the

relevance of the focus of the study to this review. Full weights of evidence for each study are given in Appendix 4.2.

Publication bias
It was not possible to draw a funnel plot in order to investigate the presence of publication bias in the in-depth review because there were too few trials included in the study: only four randomised and controlled trials out of a total of 12 studies.

Synthesis [of high/medium and medium/high studies]
The narrative overview must begin with the studies rated high and high/medium or medium/high: Elley *et al.* (1975, 1979) (high to medium); Bateman and Zidonis (1966) (medium to high); and Fogel and Ehri (2000) (high).

The two Elley *et al.* papers are based on a single study, the aim of which was to 'carry out a study of the effects of traditional and transformational grammar on children's writing skills, and in so doing to avoid the deficiencies of previous research on the subject' (Elley *et al.*, 1975, p 27). The investigation, a controlled trial, was conducted in one co-educational high school in Auckland in the early 1970s. Within the school, '248 pupils in eight matched classes of average ability were taught, observed and regularly assessed from the beginning of their third-form year…to the latter part of their fifth-form year' (op. cit., p 28), i.e. from age 13/14 to 15/16. Elley *et al.* describe how, at the outset, 'one bright and three slow-learning classes were deliberately excluded from the study, thus rendering [the sample] more homogenous, and increasing the chance of identifying systematic differences between groups' (ibid). The experimental pupils 'were classified into eight matched classes of 31 pupils' on the basis of a number of tests, and additional matching criteria were 'ethnic group, sex, contributing school, and subject options' (ibid). Although the pupils were allocated as individuals to the eight classes, the study – after this allocation – works as a cluster trial as the pupils in the eight classes were taught together. The three experimental groups were divided into 3, 3 and 2 classes, and the pupils tested during the intervention period and at the end.

The three courses studied by the three groups were, essentially, a transformational grammar course; a reading-writing course, which substituted 'extra reading and creative writing for the transformational grammar strand' (op. cit., p 29); and a traditional grammar course. In summary, the results of the study reveal that in the third year (i.e.13/14 year-olds), on none of the 12 variables chosen for analysis did any English programme show a significant superiority. The transformational grammar classes liked writing less than the other groups. In the fourth year (14/15 year-olds), only one comparison (from 30 possible) showed significant differences (on essay content). In the School Certificate Examination, there were no significant differences between the three programmes. In the fifth year (15/16 year-olds), only two of the 12 variables listed showed any significant differences (sentence-combining test and English usage test). Again, in the School Certificate Examination, there were no significant differences between the three groups. In attitude tests, transformational grammar pupils found English lessons more 'repetitive' and 'useless' than the other groups. Elley *et al.* conclude 'that English grammar, whether traditional or transformational, has virtually no influence on the language growth of typical secondary school students' (op. cit., p 38). Although they suggest that the benefits to be derived from grammar teaching might be confined to bright children, i.e. 'those who can readily manipulate and apply those syntactic structures which they studied' (op. cit., p

39), the study had deliberately omitted the top 8%-10% of pupils in the year and the suggestion remains a speculation. Nor can the results be generalised to older students, or to learners of English as a second language.

The other studies rated relatively high by the review team in relation to the specific focus of our study were those by Bateman and Zidonis (1966) and Fogel and Ehri (2000). Bateman and Zidonis examined a cohort of pupils moving from ninth to tenth grade in an American high school. Fifty pupils were assigned randomly to two sections and teachers were assigned randomly to the two classes. Although this is an individually randomised controlled trial, in that pupils were assigned randomly to the two groups, thereafter the study works as a cluster trial because the pupils in each class were taught together. The study sought to measure the effect that the teaching of a generative grammar had upon the writing of pupils, aiming also 'to help them become stylists who have expanded their capability of generating varied and well-formed sentences of the language' (Bateman and Zidonis, 1966, p ix). The experimental group was required to learn from special grammatical materials provided by the investigators. Written compositions were collected from both groups during the first three months of the first year and the last three months of the second year of the project. Results show that there was a greater increase in the average structural complexity scores for well-formed sentences in the experimental group than in the control group; and that measures of five grammatical operations indicated that the experimental group had the edge. However, the greatest changes were made by only four students, one of whom showed a very large increase in structural complexity. Similarly, the experimental group was better able to hold in check an increase in malformed sentences as structural complexity, in general, increased.

In terms of the research question we are trying to answer, a well-designed randomised controlled trial would appear to give us the most valid and reliable results. However, Bateman and Zidonis conclude, quite rightly, that the findings should be treated with caution because, although criteria of internal validity were met through 'careful randomisation procedures', the sampling requirements needed to meet criteria of external validity could not be adequately fulfilled. Furthermore, the analyses do not take the clustered nature of the data into account. Nevertheless, the study suggests, tentatively, that high school students can learn the principles of generative grammar; that a knowledge of generative grammar enables such students to increase the proportion of well-formed sentences they write; and that a knowledge of generative grammar can enable students to reduce the occurrence of errors in writing.

It is not possible to systematically synthesise the results of the Elley *et al.* and Bateman and Zidonis studies. Firstly, the transformational grammatical approach of Elley *et al.*, based as it is on materials from the Oregon Curriculum (Kitzhaber, 1968), uses – we assume - different intervention materials from the unspecified 'special grammatical materials' of Bateman and Zidonis. Secondly, the analytical framework of the two studies is different, with Elley *at al*. using 12 variables for analysis and Bateman and Zidonis, 46. Thirdly, we cannot rule out from either study, for different reasons, methodological invalidity or unreliability. Fourthly, there is insufficient detail given in Bateman and Zidonis of the intervention or of the analytical tools used (hence the lower rating than Elley *et al.* in terms of weight of evidence). Fifthly, there is no clear comparability between the two studies because Elley *et al.* use what they call a 'transformational' approach, and Bateman and Zidonis use a 'generative' approach to transformational/generative

grammar. The relationship between the two, and to transformational and generative grammars and theories, is not clearly articulated.

Fogel and Ehri's (2000) aim was 'to examine how to structure dialect instruction so that it is effective in teaching Standard English forms to students who use Black English Vernacular in their writing' (op. cit., p 215). This is thus a very different study from the previous two, discussed above, in that it does not look at transformational/generative or traditional grammar teaching. Rather, its focus is on third and fourth grade (8/9 year old) African-American pupils who exhibited Black English Vernacular (BEV) features in their written work and yet who must acquire writing competence in Standard English (SE). Three training procedures were designed to teach elementary school pupils six SE syntactic forms: the Exposure-only (E) condition, in which pupils listened to stories that included multiple instances of the SE forms; the Exposure/Strategies condition (ES), in which pupils listened and in which the teacher labelled and illustrated use of the six SE forms; and the Exposure/Strategies/Practice (ESP) condition, in which pupils added to the above procedures by translating sentences from BEV to SE and then received feedback from the teacher on their accuracy. As Fogel and Ehri describe, 'effects of the treatments were assessed with measures of students' ability to translate BEV sentences into SE sentences and their ability to use SE forms in their free writing. Students' confidence in their ability to write using "correct grammatical" or SE forms was assessed with a self-efficacy measure' (op. cit., pp 227-8).

The results of this study showed that the ESP treatment was more effective in teaching students to write using SE forms than either of the other two treatments. What the study reveals is, perhaps, more pedagogical than grammar-based: differences between BEV and SE are grammatical issues, but it is not until such differences are understood and then practised in writing, that they take effect.

In summary, Elley *et al.* conclude that English grammar teaching, whether traditional or transformational, has virtually no influence on the language growth of typical secondary school students. Bateman and Zidonis conclude, tentatively, that a generative grammar approach does make a difference to syntactic quality and to the control of malformed sentences. Because of the relative quality of the two studies, methodologically, the results of the Elley *et al.* study has a higher weight of evidence, but neither study can be said to be conclusive. Fogel and Ehri present a different kind of study in which mastery of standard English written forms is improved for elementary school African-American pupils by a process of exposure, strategies for labelling and identifying grammatical features, and, crucially, practising writing in these forms and receiving teacher feedback. However, short-term feedback is not enough to cause change in pupils of this age. As the authors point out, 'further research is needed to determine whether more extensive and repeated use of the procedures would result in increased achievement; [because] instruction was limited to six forms…it is not clear whether findings would generalize to other more complex syntactic forms [nor] whether the performance differences that were observed would be maintained over time. These remain questions for further research' (Fogel and Ehri, 2000, p 230).

Table 4.2: Effect sizes (with 95% CIs) of interventions evaluated by RCTs and CTs included in the in-depth review and of medium to high weight of evidence (WoE)

Author, date	Study type	Overall WoE	Intervention	Outcome measure	Results and effect sizes (with confidence intervals)
Bateman and Zidonis, 1966	RCT	Medium to high	Each of the two groups studied the 'regular' curriculum. The experimental group studied materials specially adapted from the area of generative grammar.	Total words in writing samples; frequencies of well-formed sentences and malformed sentences; average structural complexity scores for well-formed sentences; average structural complexity scores for malformed sentences; proportion of well-formed sentences; change in frequency of class 1 errors; change in frequency of class 2, 3 and 4 errors; change in frequency of class 5 errors; change in frequency of total errors.	No significant differences between the experimental and control classes for average structural complexity scores for well- and malformed sentences. Significant difference between the control and experimental classes for change in proportion of well-formed sentences. Significant difference between the control and experimental classes for change in frequency of class 1 errors. Significant difference between the control and experimental class for change in frequency of class 5 errors. Significant difference between experimental and control classes for change in frequency of total errors. Effect sizes: Not possible for reviewers to calculate effect sizes (no SDs).
Fogel and Ehri, 2000	RCT (cluster)	High	Third and fourth grade students received one of three treatments to increase their use of Standard English (SE) in their writing: exposure to SE features in stories; story exposure + explanation of SE rules; story exposure, SE rule instruction, and guided practice transforming sentences from Black English Vernacular to SE features.	Translation tasks (students were asked to translate five sentences written with BEV into their corresponding SE forms.	The third treatment proved the most effective in enabling students to translate BEV sentences into SE forms and to employ the targeted features in their free writing (although there is no pre-test for writing production). Effect sizes: Exposure vs. Exposure and Strategies Non-sig. +ve effect for Exposure and Strategies – 0.0867 (-0.428 to 0.60) Exposure vs. Exposure, Strategies and Practice Sig +ve effect for Exposure, Strategies and Practice – 0.948 (0.403 to 1.47) Exposure and Strategies vs. Exposure, Strategies, and Practice Sig. +ve effect for

Author, date	Study type	Overall WoE	Intervention	Outcome measure	Results and effect sizes (with confidence intervals)
					Exposure, Strategies and Practice – 0.997 (0.46 to 1.53)
Elley *et al.*, 1975, 1979	CT	High to medium	Three matched groups of students studied three different English programmes for a period of three years. Two of the groups studied the Oregon Curriculum, one with, and the other without the transformational grammar (TG) strand. The third group took a conventional English course emphasising the skills of reading, writing, listening and speaking, and providing exercises in traditional grammar concepts.	Essays, tests of language skills	The results show that the effects of a study of transformational grammar on the language growth of secondary school pupils are negligible. The pupils who studied a course containing elements of traditional grammar showed no measurable benefits. The group that studies no formal grammar demonstrated competence in writing fully equal to that shown by the two grammar groups. Effect sizes (for end of first year): Transformational Grammar vs. Reading-Writing Non-sig. +ve effect for Reading-Writing – 0.024 (-0.268) Transformational Grammar vs. Lets Learn English Non-sig. +ve effect for TG –0.09 (-0.412 to 0.236) RW vs. LLE Non-sig. +ve effect for RW - -0.122 (-0.449 to 0.205).

Other studies
Of the remaining studies in the in-depth review, four were rated medium/low in terms of overall weight of evidence (Hilfman, 1970; McNeill, 1994; Rousseau and Poulson, 1995; Stock, 1980) and three were rated low (Roberts and Boggase, 1992; Satterfield and Powers, 1996; Stone and Serwatka, 1982).

Stock (1980) looked at second grade pupils, the purpose of her study being 'to ascertain if the teaching and practicing of sentence building, by means of teaching specific capitalized and punctuated noun-verb sentence patterns, would result in the increase of these sentence patterns in the written narrative compositions of a sample of second grade children' (op. cit., p2). This is a controlled trial with prospective allocation to experimental and controlled groups, using 19 matched pairs. The experimental group was taught a total of 15 lessons over a period of six weeks. The sentence pattern used was the noun-verb (subject-predicate) pattern. Lessons dealt with nouns, verbs, articles, adjectives, adverbs, and how these elements combine into various noun-verb patterns. Results showed that, on the one hand, twice as many experimental pupils as control pupils increased the length of their T-units; on the other, teaching and having children practise a specific noun-

verb sentence pattern, complete with initial capitalisation and end punctuation, did not result in the use of the same pattern in children's writing. However, despite an appropriate study design for a question of effectiveness, the lack of detail on conceptual focus, sample, context and measures means that the results can only be given low/medium weight overall.

McNeill (1994) is a small-scale pre- and post-test study of four individual 11-16 deaf pupils in response to a single intervention designed to improve their writing skills. The aim of the study was to find out whether 'natural language instruction of certain cohesion devices would help improve the "syntactic cohesion" of the writing produced by four hearing-impaired students' (op. cit., p 89). The average gain in correct use of target conjunctions by the pupils was from 60% pre-instruction to 87% post-instruction, with the figure recorded at 83% six weeks after post-instruction, showing that progress was maintained. Content analyses showed that the pupils' papers included longer phrases and sentences, and that there was more use of connectives like 'however' and 'therefore'. Although the author concludes (albeit tentatively), that 'instructing deaf students with advanced written English skills in the use of cohesive devices in a non-English [history] class setting' (op. cit., p 94) is successful, a number of factors make this a study which is hardly generalisable. First, the sample is small. Second, the pre- and post-test design for a causal study has low internal validity, because of possible regression to the mean effects and possible selection bias. Third, there is no detailed account of the context of the primary study, nor of the sampling frame from which the four pupils are selected. This good small-scale study might have led, more usefully, to a series of refined questions for further research, rather than to tentative generalisations.

The study by Hilfman (1970) also has some limitations. Hilfman's aim was to help second grade pupils expand and elaborate their sentences by increasing their understanding of syntactic units, hoping that after exposure to word form classes and through their use of sentence expansion techniques, they would write longer and more complex sentences (in that order). The pre- and post-test design, with a mid-test used during the intervention period, was used with 18 second grade pupils and one first grade pupil. Although the author concluded that 'the study provided some evidence that some second grade children can, through instruction, expand and elaborate their sentences by increasing their understanding of syntactic units' (op. cit., p 211), there is no parallel group controlling for extraneous variables, nor any control for regression to the mean effects or temporal effects; no contextual detail; and no details of validity or reliability.

Rousseau and Poulson (1995) is a complex study, the main aim of which was to improve the quality of descriptive writing of behaviourally disordered students. Part of the problem with the study is that its title – 'Using sentence combining to teach the use of adjectives in writing to severely behaviorally disordered students' – is a misnomer. In fact, the study took the form of three 'treatments' or interventions: first, the use of a worksheet on punctuation and capitalisation and the reinforcement of the use of adjectives in story-writing; second, sentence-combining exercises to teach the correct use of adjectives; and third, a combination of sentence-combining and praise in the use of adjectives. What the researchers found in their study of three participants (Chad, aged 12; Andy, aged 9; and Joe, aged 13) was that there was an increase in the number of adjectives used per T-unit ('an independent clause including subordinate clauses attached to or embedded in it' but not conjoined with a conjunction) after the first intervention. This improvement in the correct use of adjectives was maintained, but not

improved upon, after the second and third interventions. The conclusion from the study was that 'improvement in the composition skills of academically deficient [sic] students was demonstrated as a function of reinforcement and simple instructions' (op. cit., p 15), rather than from sentence-combining or further reinforcement and praise. In other words, intervention one was effective, but two and three were not particularly so in this experiment. The study was given a medium rating because there was no clear research question or hypothesis; the baseline design was reasonable but the sample was small; there was a lack of detail about validating some of the instruments used; and the feedback/praise appears not to have been linguistic. The results cannot be said to be generalisable.

The remaining studies provide low weight of evidence for the current in-depth review. Roberts and Boggase (1992) looked at a grade 10 class with a view to examining non-intrusive grammar in writing, but with too small a sample, poorly conceived aims and research questions, the conflation of results with conclusions, and an ill-disciplined methodology, the claims that there are significant gains in fluency and a reduction in sentence-boundary errors for a number of students in the sample seem to be over-stated. There is no account in the study of how validity and reliability were addressed.

Satterfield and Powers (1996) aimed to see whether a combination of whole language and traditional approaches could improve deaf pupils' acquisition of English as evidenced through their writing. Five deaf pupils engaged in an intervention that was delivered as part of their routine teaching over six months. There was no pre- or post-test; rather, the study is a curriculum innovation from which conclusions are drawn. These are that the pupils were encouraged to write more than previously; that pupils retained grammatical concepts, making them a permanent part of their writing repertoire; and that whole language strategies, combined with mini-lessons in grammatical structure, enabled each of the five pupils to show improvement in written expression by the end of the school year. Although the focus of the study is relevant to the study – and the idea of combining whole language approaches with mini-grammar lessons is attractive pedagogically – the low level of execution of this particular study makes its evidence lightweight. The sample is small and unclearly derived; there is a lack of methodological account; and the results and conclusions are fused.

Finally, Stone and Serwatka's (1982) study of a 14-year old 'retarded' girl concludes that 'syntactic errors could be reduced through a process of oral self-patterning with the teacher serving as a guide' (p 74). The study is interesting, but there are no details of data-collection methods. Reliability was checked during each of the five (ABACA) stages of data-collection, but there is no comment on the validity of the methods. A generalisation is made on the basis of the study of a single participant.

4.4 In-depth review: quality assurance results

There are 10 studies in the in-depth review for this research question. All 10 studies were independently double data-extracted by Richard Andrews, Sue Beverton, Terry Locke, Graham Low, Carole Torgerson and Die Zhu, and by Diana Elbourne and Jo Garcia (both from the EPPI-Centre). The data-extractions were compared and all disagreements resolved. The English Review Group data-

extraction for each of the 10 studies was then uploaded to the EPPI-Centre's Research Evidence in Education Library (REEL).

5. FINDINGS AND IMPLICATIONS

5.1 Summary of principal findings

The results of the present in-depth review point to one clear conclusion: that there is no high quality evidence to counter the prevailing belief that the teaching of the principles underlying and informing word order or 'syntax' has virtually no influence on the writing quality or accuracy of 5 to 16 year-olds. This conclusion remains the case whether the grammar teaching is based on the 'traditional' approach of emphasising word order and parts of speech; or on the 'transformational' approach, which is based on generative-transformational grammar.

Nearly all our included studies were experimental (i.e. researcher-manipulated as opposed to naturally-occurring evaluations), a highly appropriate design for testing causality.

The most authoritative study we examined – that by Elley *et al.* (1975, 1979) – concluded that neither type of grammar teaching of syntax (traditional versus transformational), had a significant influence on the writing development of typical secondary school students. Bateman and Zidonis (1996), a study of ninth and tenth grade pupils, concluded that a generative grammar approach did have a positive effect on syntactic quality and the control of badly formed sentences, but their study was of a slightly lower quality than Elley *et al.*, and thus cannot be afforded as much weight. The third paper that we rated as having considerable weight of evidence was that by Fogel and Ehri (2000), which focused on elementary/ primary school African-American pupils. This paper suggested that, by using a range of methods to teach writing at an early stage, including the actual practising of writing, standard English written forms could be mastered. However, the short-term results were not followed up and the authors suggest that further research is needed to determine whether extensive and repeated use of procedures used would result in increased and sustained achievement.

It was not possible to calculate effect sizes in the Bateman and Zidonis study, as there were no standard deviations reported. The experimental group in the Fogel and Ehri study showed marginal positive effects for exposure, strategies and practice in the light of the intervention, with one of the treatments proving the most effective in enabling students to translate Black English Vernacular sentences into Standard English forms and to employ the targeted features in their free writing. The results of the Elley *et al.* study show that the effects of the teaching of transformational grammar are negligible, and those of traditional grammar non-existent. Those who were taught no formal grammar demonstrated competence in writing fully equal to that of the two experimental groups.

Seven other papers were examined in the in-depth review. Because none of these were rated as above medium in overall weight of evidence, they have not been used to determine the overall findings of the review. They are discussed in detail in the body of the report itself.

When we compare the results of the present systematic review with one previous systematic review and 15 previous non-systematic reviews published from 1997 to 2001, we find that ten of the reviews (Abrahamson, 1977; Bamberg, 1981; Crowhurst, 1980; Elley, 1994; Gann, 1984; Hillocks, 1984, 1986; Tomlinson, 1994; Ulin and Schlerman, 1978; Walsh, 1991; Wyse, 2001) declare that the teaching of formal grammar has no effect on the quality or accuracy of writing; four come to no conclusion; one (Hudson, 2000) declares that 'the idea that grammar teaching improves children's writing skills is much better supported by the available research than is commonly supposed' (p 4). However, the review itself does not look in detail at all the available studies and shows that traditional grammar teaching is ineffective. The previous systematic review finds that sentence-combining 'on the average…[is]…more than twice as effective as free writing as a means of enhancing the quality of student writing' (Hillocks, 1984, p 161). We pursue the question of sentence-combining in another systematic review.

5.2 Strengths and limitations of this systematic review

The strengths of the review are several in number. First, it is the most comprehensive review so far undertaken on the effect of grammar teaching (syntax) on writing quality and accuracy. Over 4,500 studies were screened, initially. Second, the distillation process that enabled us to arrive at a small number of studies that were of high quality, and a slightly larger number that were highly relevant to answering our research question, was an exhaustive and entirely transparent one. By using EPPI-Centre methodology, we are confident that we have done justice to the published and unpublished (but in the public domain) research, while excluding doctoral and masters theses on the topic. Third, the systematic review has not only been of primary research on the topic, but also of previous reviews of research in the field.

There are limitations, as in all research. First, our focus is sharply on effectiveness and does not consider wider issues of impact or influence. Second, we limited our searches to published work and unpublished work in the public domain, but we have excluded unpublished dissertations. (We do not consider this a serious limitation, as the review spans over a hundred years and we would expect ground-breaking doctoral research to have been published. Where it is not, we have addressed this in the background sections.) Third, the quality of seven out of ten of the primary studies included in the in-depth review is a limitation on the review, as is the lack of recent research in the effect of teaching grammar (syntax) on writing. We did not search for studies on the subject not written in English; this was not thought to be a serious limitation. Finally, we cannot be absolutely sure that we have not missed some key published research, despite our exhaustive attempts to be comprehensive. In this latter regard, we look forward to hearing from correspondents who might be able to point us towards further research that we have, inadvertently, overlooked.

5.3 Implications

There are three main implications of the systematic review and they can be categorised under implications for practice, for policy and for further research.

In terms of practice, the main implication of our findings is that the evidence base to justify the teaching of syntax in order to improve the quality and accuracy of writing whether traditional or generative/transformational, is very small. This is not to say that the teaching of such grammar might not be of value in itself, or that it might lead to enhanced knowledge and awareness of how language works, and of systems of language use. But the clear implication, based on the small number of studies providing high quality research evidence, is that the teaching of syntax in English to 5 to 16 year-olds in order to improve writing, should cease to be part of the curriculum unless in the context of rigorous evaluative research.

It was not our brief in the present review to suggest what does work in improving the quality and accuracy of writing in English for 5 to 16 year-olds, but the implication is that, if formal grammar teaching does not work, then practices based on theories such as 'you learn to write by writing' need to be given more credence and subjected themselves to further systematic review. Whether there is space in the curriculum to teach grammar for its own sake, or for other purposes, remains to be seen.

Finally, the implications for further research are various. Despite a hundred years of concern about the issue of the teaching of grammar and thousands of research studies, the high-quality research base for claiming the efficacy of grammar teaching is small. The first implication, then, is that there should be a conclusive, large-scale and well-designed randomised controlled trial to answer the question about whether grammar teaching (of different kinds not researched to date) does improve the writing quality and accuracy of 5 to 16 year-olds. Such a study should have a longitudinal dimension to test whether any significant effects are sustained.

While we do not claim the final word on the question, the present review has been the largest systematic review in the history of research on the topic to date. This does not mean that other reviews of different aspects of the question of the relationship between grammar teaching and writing quality and accuracy cannot be undertaken. The specific focus of this review has been on the teaching of syntax and a complementary review we are undertaking is on sentence-combining, both coming under the umbrella of 'grammar' teaching.

We hope to have established a landmark in studies on the effectiveness of grammar teaching in the development of writing quality and accuracy in school-age children. If this is a landmark, it points the way to further research in the field, where the territory of debate will be somewhat different. We now know that there is no high quality evidence that the teaching of formal or generative/transformational grammar is effective with regard to writing development. Having established that much, we can now go on to research what is effective, and to ask clearer and more pertinent questions about what works in the development of young people's literacy.

6. REFERENCES

6.1 Studies included in map and synthesis

Reviews
Abrahamson RF (1977) *The Effects of Formal Grammar Instruction vs the Effects of Sentence Combining Instruction on Student Writing: A Collection of Evaluative Abstracts of Pertinent Research Documents*. Houston, TX, USA: University of Houston.

Amiran E, Mann J (1982) *Written Composition, Grades K-12: Literature Synthesis and Report.* Portland, OR, USA: Northwest Regional Educational Laboratory. Educational Resources Information Center (ERIC) document number ED213034.

Bamberg B (1981) *Making Research Work for the Composition Teacher*. California: University of Southern California. ERIC document number ED208408.

Crowhurst M (1980) *The Effect of Syntactic Complexity on Writing Quality: A Review of Research*. British Columbia: University of British Columbia. ERIC document number ED202024.

Elley WB (1994) Grammar teaching and language skill. In: Asher R (ed.) *Encyclopedia of Language and Linguistics*. Oxford: Pergamon, pages 1468-1471.

Gann M (1984) Teaching grammar: is structural linguistics really better? *English Quarterly* **17:** 31-53.

Hillocks G Jr (1984) What works in teaching composition: a meta-analysis of experimental treatment studies. *American Journal of Education* **93:** 133-170.

Hillocks G Jr (1986) *Research on Written Composition: New Directions for Teaching*. Urbana, IL, USA: National Conference on Research in English. ERIC document number ED265552.

Hudson R (2000) *Grammar Teaching and Writing Skills: The Research Evidence*. London: University College London, http://www.phon.ucl.ac.uk/home/dick/bib.htm.

Kolln M (1996) Rhetorical grammar: a modification lesson. *English Journal* **85:** 25-31.

Lawlor J (1980) *Improving Student Writing through Sentence Combining: A Literature Review. Technical Note*. Los Alamitos, CA, USA: Southwest Regional Laboratory for Educational Research and Development. ERIC document number ED192356.

Matzen RN Jr, Zhang W-S, Hartwell P (1995) *The Role of Traditional Grammar Instruction in the Teaching of Writing: A Selected, Annotated Bibliography*.

Indiana, PA, USA: Indiana University of Pennsylvania. ERIC document number ED396328.

Newkirk T (1978) Grammar instruction and writing: what we don't know. Close-up: grammar and composition. *English Journal* **67:** 46-48.

Ney JW (1980) A short history of sentence combining: its limitations and use. *English Education* **11:** 169-177.

Phillips SE (1996) *Sentence Combining: A Literature Review*. Dallas, TX, USA: University of Texas. ERIC document number ED398589.

Qualifications and Curriculum Authority (QCA) (1998) *Recent Research on Grammar Teaching*. Hayes, Middlesex: QCA Publications.

Seidenberg PL (1989) Relating text-processing research to reading and writing instruction for learning disabled students. *Learning Disabilities Focus* **5:** 4-12.

Sternglass MS (1979) *Creating the Memory of Unheard Sentences*. Indiana, PA, USA: Indiana University of Pennsylvania. ERIC document number ED176258.

Stewart MF (1979) Sentence-combining: its past, present and future. *English Quarterly* **12:** 21-36.

Stotsky SL (1975) Sentence-combining as a curricular activity: its effect on written language development and reading comprehension. *Research in the Teaching of English* **9:** 30-71.

Tomlinson D (1994) Errors in the research into the effectiveness of grammar teaching. *English in Education* **28:** 20-26.

Ulin RO, Schlerman BJ (1978) For the sake of teaching writing, deliver us from traditional grammar. *High School Journal* **62:** 58-68.

Walsh SM (1991) *Breakthroughs in Composition Instruction Methods Without Evidence of Tangible Improvements in Students' Composition: When Will Change Come?* New York, NY, USA: State University of New York. ERIC document number ED336744.

White RS, Karl H (1980) Reading, writing and sentence combining: the track record. *Reading Improvement* **17:** 226-232.

Wyse D (2001) Grammar. For writing? A critical review of empirical evidence. *British Journal of Educational Studies* **49:** 411-427.

Primary studies
Bateman DR, Zidonis FJ (1966) *The Effect of A Study of Transformational Grammar on the Writing of Ninth and Tenth Graders*. Champagne, IL, USA: National Council of Teachers of English.

Calkins LM (1979) *When Children Want to Punctuate: Basic Skills Belong in Context*. Durham, NH, USA: University of New Hampshire. ERIC document number ED170766.

Combs WE (1976) Further effects of sentence-combining practice on writing ability. *Research in the Teaching of English* **10:** 137-149.

Combs WE (1977) Sentence-combining practice: do gains in judgments of writing 'quality' persist? *Journal of Educational Research* **70:** 318-321.

Combs WE, Wilhelmsen K (1979) In-class 'action' research benefits research, teacher and students. *High School Journal* **62:** 267-271.

Elley WB, Barham IH, Lamb H, Wyllie M (1975) The role of grammar in a secondary school curriculum. *New Zealand Council for Educational Studies* **10:** 26-41.

Elley WB, Barham IH, Lamb H, Wyllie M (1979) *The Role of Grammar in a Secondary School Curriculum. Educational Research Series No 60*. Wellington: New Zealand Council for Educational Research. ERIC document number ED185588.

Fogel H, Ehri LC (2000) Teaching elementary students who speak black English vernacular to write in standard English: effects of dialect transformation practice. *Contemporary Educational Psychology* **25:** 212-235.

Gordon CJ (1990) Contexts for expository text structure use. *Reading Research and Instruction* **29:** 55-72.

Hilfman T (1970) Can second grade children write more complex sentences? *Elementary English* **47:** 209-214.

Hunt KW, O'Donnell R (1970) *An Elementary School Curriculum to Develop Better Writing Skills*. Tallahassee, FL, USA: Florida State University.

MacNeill TB (1982) *The Effect of Sentence-Combining Practice on the Development of Reading Comprehension and the Written Syntactic Skills of Ninth Grade Students*. Alberta: University of Alberta. ERIC document number ED217415.

McAfee D (1981) *Effect of Sentence Combining on Fifth Grade Reading and Writing Achievement*. Dallas, TX, USA: University of Texas. ERIC document number ED217388.

McNeill JH (1994) Instruction for deaf students in syntactic cohesion. *Acehi Journal/Revue Aceda* **20:** 88-95.

Mellon J (1969) *Transformational Sentence Combining. Research Report No. 10*. Urbana, IL, USA: National Council of Teachers of English.

Melvin MP (1980) *The Effects Of Sentence Combining Instruction on Syntactic Maturity, Reading Achievement, and Language Arts Skills Achievement*. Oxford, OH, USA: Miami University. ERIC document ED191007.

Miller BD, Ney JW (1967) Oral drills and writing improvement in the fourth grade. *Journal of Experimental Education* **36:** 93-99.

Miller BD, Ney JW (1968) The effect of systematic oral exercises on the writing of fourth-grade students. *Research in the Teaching of English* **2:** 44-61.

Ney JW (1976) *Sentence Combining and Reading*. Tempe, AZ, USA: Arizona State University. ERIC document number ED161080.

Nutter N, Safran SP (1983) *Sentence Combining and the Learning Disabled Student*. Athens, OH, USA: Ohio University. ERIC document number ED252994.

O'Hare F (1973) *Sentence Combining: Improving Student Writing without Formal Grammar Instruction. Research Report No 15.* Urbana, IL, USA: National Council of Teachers of English.

Pedersen EL (1978) *Sentence-combining Practice: Training that Improves Student Writing*. Utah: Weber State College. ERIC document number ED169567.

Roberts CM, Boggase BA (1992) Non-intrusive grammar in writing. Paper presented to the Annual Conference on Computers and Writing. Indianapolis, USA: May 1-3. ERIC document number ED348684.

Robinson CF (1978) *An Investigation of New Rhetoric Lessons for Improved Written Composition on the Secondary School Level*. Hartford, CT, USA: Hartford Public School System. ERIC document number ED188196.

Rousseau MK (1989) *Increasing the Use of Compound Predicates in the Written Compositions of Students with Mild Learning Handicaps*. New York: City University of New York. ERIC document number ED342154.

Rousseau MK, Poulson CL (1985) *Using Sentence-combining to Teach the use of Adjectives in Writing to Severely Behaviorally Disordered Students*. New York: City University of New York. ERIC document number ED342153.

Satterfield J, Powers A (1996) Write on! Journals open to success. *Perspectives in Education and Deafness* **15:** 2-5.

Stock R (1980) *The Effect of Teaching Sentence Patterns on the Written Sentence Structures of Grade Two Children*. Manitoba, Canada. ERIC document number ED208414.

Stoddard EP, Renzulli JS (1983) Improving the writing skills of talent poor students. *Gifted Child Quarterly* **27:** 21-27.

Stone AK, Serwatka TS (1982) Reducing syntactic errors in written responses of a retarded adolescent through oral patterning. *Education and Training in Mental Retardation and Developmental Disabilities* **17:** 71-74.

Vitale MR, King FJ, Shontz DW, Huntley GM (1971) Effect of sentence-combining exercises upon several restricted written composition tasks. *Journal of Educational Psychology* **62:** 521-525.

Welch M (1992) The PLEASE strategy: a metacognitive learning strategy for improving the paragraph writing of students with mild learning disabilities. *Learning Disability Quarterly* **15:** 119-128.

Willig JB (1985) *Using Instruction in Sentence Combining to Improve Syntactic Maturity in Student Writing*. Florida: Nova University. ERIC document number ED323557.

6.2 Other references used in the text of the report

Asker W (1923) Does knowledge of formal grammar function? *School and Society* (27 January): 109-111.

Benfer M (1935) Sentence sense in relation to subject and predicate. Unpublished Master's thesis. University of Iowa.

Boraas J (1917) Formal grammar and the practical mastery of English. Unpublished doctoral thesis. University of Minnesota.

Braddock R, Lloyd-Jones R and Schoer L (1963) *Research on Written Composition*. Illinois, USA: National Council of Teachers of English.

Carter R (ed.) (1990) *Knowledge about Language and the Curriculum: The LINC Reader.* London: Hodder and Stoughton.

Catherwood C (1932) A study of the relationship between a knowledge of rules and ability to correct grammatical errors and between identification of sentences and knowledge of subject and predicate. Master's thesis. University of Minnesota.

Chomsky N (1957) *Syntactic Structures.* The Hague: Mouton.

Damasio A (2000) *The Feeling of What Happens: Body, Emotion and the Making of Consciousness.* London: Vintage.

Department of Education and Science (DES) (1975) *A Language for Life* (The Bullock Report). London: Her Majesty's Stationery Office (HMSO).

DES (1988) *Report of the Committee of Inquiry into the Teaching of the English Language* (The Kingman Report). London: HMSO.

Department for Education and Employment (DfEE) (2000) *The National Literacy Strategy: Grammar for Writing*. London: DfEE (book and video).

DfEE/QCA (1999) *The National Curriculum for England: English.* London: DfEE/QCA.

Derewianka B (1998) *A Grammar Companion for Primary Teachers.* Newtown, NSW: Primary English Teaching Association.

Dixon RMW (1965) *What is Language? A New Approach to Linguistic Description.* London: Longmans, Green and Co Ltd.

EPPI-Centre (2002a) *Core Keywording Strategy: Data Collection for a Register of Educational Research.* (Version 0.9.7). London: EPPI-Centre, Social Science Research Unit.

EPPI-Centre (2002b) *EPPI-Reviewer®.* (Version 0.9.7). London: EPPI-Centre, Social Science Research Unit.

EPPI-Centre (2002c) *Guidelines for Extracting Data and Quality Assessing Primary Studies in Educational Research*. (Version 0.9.7). London: EPPI-Centre, Social Science Research Unit.

Fairclough N (1992) *Discourse and Social Change.* Cambridge: Polity Press.

Firth JR (1935) The technique of semantics. Transactions of the Philological Society. Reprinted in: Firth JR *Papers in Linguistics 1934-1951.* (1957). London: Oxford University Press (OUP).

Halliday MAK, Hasan R (1985) *Language, Context and Text: Aspects of Language In a Social Semiotic Perspective.* Geelong, VIC, Australia: Deakin University Press.

Harris RJ (1962) An experimental enquiry into the functions and value of formal grammar in the teaching of English, with special reference to the teaching of correct written English to children aged twelve to fourteen. Unpublished PhD thesis. University of London.

Hodge R, Kress G (1993) *Language as Ideology*. 2nd edition. London: Routledge.

Hudson R (1992) *Teaching Grammar: A Guide for the National Curriculum.* Oxford: Basil Blackwell.

Hunt KW (1966) Recent measures in syntactic development. *Elementary English* **XLIII:** 732-739.

Kitzhaber AR (ed.) (1968) *The Oregon Curriculum: A Sequential Program in English.* New York: Holt, Rinehart and Winston.

Kress G (1994) *Learning to Write.* 2nd edition. London: Routledge.

Macaulay WJ (1947) The difficulty of grammar. *British Journal of Educational Psychology* **17:** 153-162.

Mandler J, Johnson NS (1977) Remembrance of things parsed: story structure and recall. *Cognitive Psychology* **9:** 111-151.

Ministry of Education (1996) *Exploring Language: A Handbook for Teachers.* Wellington, NZ: Learning Media.

Norris JM, Ortega L (2000) Effectiveness of L2 instruction: a research synthesis and quantitative meta-analysis. *Language Learning* **50:** 417-528.

Office for Standards in Education (Ofsted) (2002) *The Key Stage 3 Strategy: Evaluation of the First Year of the Pilot*. London: Ofsted.

Perera K (1984) *Children's Writing and Reading: Analysing Classroom Language.* Oxford: Basil Blackwell.

Pinker S (1995) *The Language Instinct: The New Science of Language and Mind.* London: Penguin.

Propp V (1968) *Morphology of the Folk Tale*. 2nd edition. Austin, TX, USA: University of Texas Press.

QCA (1999) *Not Whether but How: Teaching Grammar in English at Key Stages 3 and 4*. London: QCA.

Quirk R, Greenbaum S, Leech G, Svartvik J (1985) *A Comprehensive Grammar of the English Language.* London: Longman.

Rice JM (1903) Educational research: the results of a test in language and English. *Forum* **XXXV:** 209-293, 440-457.

Robinson N (1960) The relationship between knowledge of English grammar and ability in English composition. *British Journal of Educational Psychology* **30:** 184-186.

Rumelhart DE (1975) Notes on a schema for stories. In: Bobrow DG, Collins A (eds) *Representation and Understanding: Studies in Cognitive Science.* New York: Academic Press.

Saussure F de (1961) *Course in General Linguistics.* Bally C, Sechehaye A (eds) Translated by Baskin W. London: McGraw Hill.

Segal D, Barr NR (1926) Relation and achievement in formal grammar to achievement in applied grammar. *Educational Research* **14:** 401-402.

Symonds PM (1931) Practice vs. grammar in the learning of correct usage. *Journal of Educational Psychology* **22:** 81-96.

Walmsley J (1984) *The Uselessness of Formal Grammar?* Birmingham: British Association for Applied Linguistics/Linguistics Association of Great Britain, Committee for Linguistics in Education (CLIE Working Papers, number 2).

Wilkinson A (1971) *The Foundations of Language: Talking and Reading in Young People.* London: OUP.

Wright E (1989) Grammar in English teaching: scientific evidence old and new. *Quest* **38:** 11-15 (part one) and *Quest* **39:** 12-16 (part two).

APPENDIX 2.1: Inclusion and exclusion criteria

For a paper to be included in the systematic map, it will have to be a study looking at the effect of grammar teaching in English on 5 to 16 year-olds' accuracy and quality in written composition. As the focus of the study is on the *effects* of grammar teaching, papers using methods to identify any such effects are required. This implies the following study types, classified according to the EPPI-Centre taxonomy of study type contained in its core keywording strategy (EPPI-Centre 2002a):

B: Exploration of relationships
C: Evaluation (naturally-occurring or researcher-manipulated)
E: Review (systematic or other review) containing at least one study exploring
   relationships or one evaluation

Inclusion criteria
- Must be a study of the effects of grammar teaching on writing
- Must focus exclusively on children and young people aged 5 to 16
- Must be in a mainstream school setting
- Must be one of the following study types: B (exploration of relationships);
  C (evaluation); E (review)
- Must be published or unpublished (but in the public domain) between 1900
  and the present
- Must be teaching of English grammar in an English-speaking country
- Must be teaching of English as first language, not foreign or second or
  additional language

Exclusion criteria
EXCLUSION ON SCOPE
One: Not grammar teaching
Two: Not children or young people aged between 5 and 16
Three: Not effects of grammar teaching on writing
Three (a)   Not teaching of English grammar (syntax) in an English-speaking
                  country

EXCLUSION ON STUDY TYPE
Four: (a) A (description)
                  (b) D (methodology)
                  (c) Editorial, commentary, book review
                  (d) Policy document
                  (e) Resource, textbook
                  (f) Bibliography
                  (g) Dissertation abstract
                  (h) Theoretical paper
                  (i) Position paper

EXCLUSION ON SETTING IN WHICH STUDY WAS CARRIED OUT

Five: English as a foreign, second or additional language (L2, EFL,
                  ESL, EAL)

Six: Not mainstream school setting

Seven: The effects of grammar teaching on the writing of pupils in a
            foreign language (e.g. Hebrew, Dutch)

APPENDIX 2.2: Search strategy for electronic databases

ERIC
1. exp *grammar/ or exp *syntax/
2. exp *sentence structure/
3. *writing (composition)/
4. *metalinguistics/
5. *cohesion (written composition)/ or *generative grammar/ or *sentence combining/ or *sentence diagraming/ or *structural grammar/ or *text structure/ or *traditional grammar/
6. *case (grammar)/ or *grammatical acceptability/
7. *transformational generative grammar/
8. *coherence/ or *paragraph composition/
9. "KAL".mp
10. 1 or 2 or 3 or 4 or 5 or 6 or 6 or 8 or 9
11. limit 10 to English language
12. limit 11 to (elementary secondary education or elementary education or primary education or intermediate grades or secondary education or middle schools or junior high schools or high schools or high school equivalency programs)
13. limit 12 to (books or conference proceedings or dissertations or "evaluative or feasibility reports" or general reports or information analyses or journal articles or "research or technical reports" or "speeches or conference papers"

PsycINFO
1. ("grammar-" in DE) or ("transformational-generative grammar" in DE)
2. "syntax-" in DE
3. "sentence-structure" in DE
4. "text-structure" in DE
5. (writ*) and (composition*)
6. "metalinguistics-" in DE
7. 1 or 2 or 3 or 4 or 5 or 6
8. limit 7 to ((AG:PY = Adolescence) or (AG:PY = childhood) or (AG:PY = school age)) and (LA:PY = English) and ((PT:PY = case-study) or (PT:PY = conference-proceedings-symposia) or (PT:PY = empirical-study) or (PT:PY = followup-study) or (PT:PY = journal-abstract) or (PT:PY = journal information) or (PT:PY = journal-review-book) or (PT:PY = literature-review-research-review) or (PT:PY = meta-analysis) or (PT:PY = prospective-study) or (PT:PY = retrospective-study) or (PT:PY = treatment-outcome-study))

SSCI
(((gramma* or synta* or sentence structure or metlinguistic* or knowledge about language or KAL))) and ((writ* or composition*)) and (child* or adolescen* or school* or education*)
Doc type = all document types
Language = English

APPENDIX 2.3: EPPI-Centre core keywords

1. Identification of report
Citation
Contact
Handsearch
Unknown
Electronic database
(Please specify.) …………………………

2. Status
Published
In press
Unpublished

3. Linked reports
Is this report linked to one or more other reports in such a way that they also report the same study?

Not linked
Linked (Please provide bibliographical details and/or unique identifier.)
………………………………………………
………………………………………………
………………………………………………
………
………………………………………………
…

4. Language (Please specify.)

……………………………………………
…

5. In which country/countries was the study carried out?
(Please specify.)

……………………………………………
……………………………………………

6. What is/are the topic focus/foci of the study?
Assessment
Classroom management
Curriculum
Equal opportunities
Methodology
Organisation and management
Policy
Teacher careers
Teaching and learning
Other (Please specify.)…………………

7 Curriculum
Art
Business studies
Citizenship
Cross-curricular
Design and technology
Environment
General
Geography
Hidden
History
ICT
Literacy – first language
Literacy further languages
Literature
Maths
Music
PSE
Phys. Ed.
Religious Ed.
Science
Vocational
Other (Please specify.) ………………

8. Programme name (Please specify.)

…………………………………………

9. What is/are the population focus/foci of the study?
Learners*
Senior management
Teaching staff
Non-teaching staff
Other education practitioners
Government
Local education authority officers
Parents
Governors
Other (Please specify.) ……………………………

10. Age of learners (years)
0-4
5-10
11-16
17-20
21 and over

11. Sex of learners
Female only
Male only
Mixed sex

12. What is/are the educational setting(s) of the study?
Community centre
Correctional institution
Government department
Higher education institution
Home
Independent school
Local education authority
Nursery school
Post-compulsory education institution
Primary school
Pupil referral unit
Residential school
Secondary school
Special needs school
Workplace
Other educational setting …………………………..

13. Which type(s) of study does this report describe?

A. Description
B. Exploration of relationships
C. Evaluation
    a. Naturally-occurring
    b. Researcher-manipulated
D. Development of methodology
E. Review
    a. Systematic review
    b. Other review
*see 14.

14. To assist with the development of a trials register please state if a researcher- manipulated evaluation is one of the following:
a. Controlled trial (non-randomised)
b. Randomised controlled trial (RCT)

Please state here if keywords have not been applied from any particular category (1-10) and the reason why (e.g. no information provided in the text).

………………………………………
………………………………………
………………………………………
………………………………………
………………………………………
………………………………………

APPENDIX 2.4: Review-specific keywords

1. On what 'type' of grammar teaching does the study focus?	2. If 'text' level, is the focus on:	3. If 'sentence' level, is the focus on:	4. What 'type' of intervention does the study involve?
a. 'text' level grammar teaching b. 'sentence' level grammar teaching	a. text structure? b. cohesion? c. coherence? d. paragraph composition? e. not applicable	a. syntax? b. sentence-diagramming? c. sentence-combining? d. punctuation? e. not applicable?	a. contextualised grammar teaching b. de-contextualised grammar teaching
5. On what kind of grammar does the study focus?	6. What 'type' of written outcomes are reported?	7. What measurements are reported?	8. What are the specific characteristics of the learners?
a. language-awareness b. meta-language c. traditional grammar d. transformative/generative grammar e. 'functional' grammar f. 'pedagogic' grammar	a. accuracy of writing (please specify) b. quality of writing (please specify)	a. test results (please specify) b. examination results (please specify) c. written composition (please specify) d. other (please specify)	a. learning difficulties b. specific learning difficulties (dyslexia) c. other (please specify) d. not applicable

For definitions, see Glossary.

APPENDIX 4.1: Summary tables for studies included in the in-depth review

Bateman DR, Zidonis FJ (1966) The effect of a study of transformational grammar on the writing of ninth and tenth graders. National Council of Teachers of English Research Report No 6.	
Country of study	USA - Ohio
Age of learners	11-16: Ninth and tenth graders A cohort moving from the ninth grade to the tenth grade over the period of study
Type of study	**Researcher-manipulated evaluation: RCT (individual)** The 50 pupils in ninth grade were 'assigned randomly' to two sections. Teachers from the Language Arts Area of the University School were assigned randomly to the two classes. Works as a cluster trial as all children in each class taught together.
Aims of study	• To measure the effect that the teaching of a 'generative grammar' (a representation of the psychological process of producing sentences) has upon the writing of ninth and tenth graders • To improve secondary students' writing and 'to help them become stylists who have expanded their capability of generating varied and well-formed sentences of the language' (p ix)
Summary of study design, including details of sample	50 pupils from ninth grade randomly allocated to experimental group or control group. Two teachers randomly assigned to the two classes. The experimental class was required to learn the special grammatical materials provided by the investigators. Written compositions were collected from both groups during the first three months of the first year and the last three months of the second year of the project (total of 12 pieces of writing - six in initial period and six in end period).
Data-collection instruments, including details of checks on reliability and validity	• *Very little detail given; 'all writing was produced by the pupils as part of class assignments'.* • Both teachers met regularly to standardise writing assignments. • *No details of reliability or validity except that the investigators developed an analytical instrument that would 'objectively assess'.* • *No external validation.*
Methods used to analyse data, including details of checks on reliability and validity	• 'The analysis of the prose consisted of a reconstruction of the transformational history of each sentence.' The analysts made decisions about whether sentences were 'well-formed' or 'malformed'. They then calculated summaries of frequencies of words, sentences, and transformations of both types of sentences (Tables 1 and 2). • Each sentence was assigned a 'structural complexity' score that represented the total number of grammatical operations it took to produce the sentence (Tables 3 and 5). • The investigators developed an analytical instrument that would objectively assess the grammatical quality of the sentences in the sample. Forty-six transformational rules served to identify the grammatical operations that each sentence in the sample reflected. The use of these grammatical structures in writing was measured, not the pupils' awareness of or ability to recall their labels. • Statistical evaluation of change in average structural complexity scores for well-formed sentences and malformed sentences by analysis of variance. • No details of reliability or validity regarding use of two reviewers or examination of negative cases. • Issues of internal/external consistency.

Summary of results	• 'The increase in average structural complexity scores for well-formed sentences was 3.793 for the control class and 9.315 for the experimental class. A difference of over five grammatical operations seems to indicate that the experimental class had significantly extended its capacity for producing complex well-formed sentences. However, a closer examination reveals that the greatest changes were made by only four students, one of whom shows a structural complexity increase of 64.741.' • 'The increase in average structural complexity scores for malformed sentences was 7.511 for the control class and 3.585 for the experimental class. Table 5 indicates that, as both classes learned to write well-formed sentences of increased complexity, they also increased the average structural complexity of malformed sentences, though the experimental class was better able to hold this tendency in check.' • Table 7: for the experimental class, there was an increase of 31.8% of well-formed sentences; for the control class, there was an increase of 3.5%. • Average reduction of class 1 errors for the experimental class was 79.06%; average reduction of class 1 errors for the control class was 39.82%.
Conclusions	The authors conclude that the findings should be treated with caution ('tentatively') because, although criteria of internal validity were adequately met through 'careful randomization procedures'), the sampling requirements needed to meet criteria of external validity could not be adequately fulfilled. • High-school students can learn the principles of generative grammar. • A knowledge of generative grammar enables students to increase the proportion of well-formed sentences they write. • There is a suggestion of a relationship between a knowledge of generative grammar and an ability to produce well-formed sentences of greater structural complexity. • Almost half the sentences written by the ninth graders were malformed. • A knowledge of generative grammar can enable students to reduce the occurrence of errors in their writing.
Weight of evidence A (trustworthiness in relation to study questions)	**Medium**
Weight of evidence B (appropriateness of research design and analysis)	**High:** Design (randomised controlled trial) highly appropriate study design for addressing question of effectiveness.
Weight of evidence C (relevance of focus of study to review)	**Medium**
Weight of evidence D (overall weight of evidence)	**Medium to high**

1. **Elley WB, Barham IH, Lamb H, Wyllie M (1975) The role of grammar in a secondary school curriculum. New Zealand council for Educational Studies, 10, 26-41.**
2. **Elley WB, Barham IH, Lamb H, Wyllie M (1979) The role of grammar in a secondary school curriculum. Educational research series no 60. Wellington: New Zealand Council for Educational Research.**

Country of study	Auckland City, NZ
Age of learners	11-16
Type of study	Researcher-manipulated evaluation: controlled trial
Aims of study	' To carry out a study of the direct effects of traditional and transformational grammar on children's writing skills, and in so doing to avoid the main deficiencies of the previous research on the subject' (Elley, 1975, p 27)
Summary of study design, including details of sample	248 pupils were allocated (after matching) into eight classes. There were three experimental groups of 3, 3 and 2 classes. Although pupils were allocated as individuals to the eight classes, thereafter this works as a cluster trial as the pupils in the eight classes were taught together. The pupils were tested during the intervention period and at the end.
Data-collection instruments, including details of checks on reliability and validity	Data-collection instruments: • Essays: a series of set essays on a variety of topics. • Tests of language skills: PAT Reading Comprehension and Vocabulary Tests (NZCER, 1969) supplemented in the last year by additional items to increase the difficulty levels. • Test of sentence-combining. • In third year, tests in spelling and listening comprehension. • Attitudes to English questionnaires. • The tests were administered at the end of the school year (in November). Precise details of the test conditions are not reported. Details of reliability: • For essays: at least two markers used, working independently. • For tests: use of standardised test. • The only reliability data reported are the use of four markers for each of the four F3 essays, and the fact that each marker read each script four times. • A reliability check on the F3 results led to three essays and two markers being used for F4 and F5. Details of validity: • For essays, none given; for language tests, use of standardised test • Free writing was measured by more than one essay test. • A range of different skills was tested.
Methods used to analyse data, including details of checks on reliability and validity	• Means, SDs and numbers in three experimental groups for end of year tests (third, fourth and fifth years). • Tests of significance - analyses of variance. • One-way and two-way ANOVA, with individual post-hoc tests. Exploratory factor analysis re the Semantic differential (for the

	attitude) scores. • Gain scores are reported at the end, but no method of analysis is reported. • No details of reliability are given. • No details of validity are given, but use of ANOVA is appropriate. • The data are analysed for gender effects and 'ability' effects.
Summary of results	• Third year: On none of the twelve variables did any English programme show a significant superiority. The TG pupils liked writing less than the other groups. • Fourth year: Only one comparison (out of 30 possible) significant. Attitude results showed no differences. • Fifth year: Only two out of 12 variables listed showed significant differences. In School Certificate Examination no significant differences between the three groups. In attitude tests, TG pupils found English more 'repetitive' and 'useless' than the other groups.
Conclusions	The investigation to determine the direct effects of a study of TG on the language growth of secondary pupils shows negligible effects. Similarly, those pupils who study a course containing elements of traditional grammar showed no measurable benefits.
Weight of evidence A (trustworthiness in relation to study questions)	**High**
Weight of evidence B (appropriateness of research design and analysis)	**High to medium:** No use of random allocation. The sample may have been from one school and two of the groups may have used an overseas programme, but generally the design allowed meaningful comparisons to be made about the effects of TG and traditional grammar.
Weight of evidence C (relevance of focus of study to review)	**Medium**
Weight of evidence D (overall weight of evidence)	**High to medium**

Fogel H and Ehri LC (2000) Teaching elementary students who speak black English vernacular to write in standard English: Effects of dialect transformation practice. *Contemporary Educational Psychology* 25: 212-235.	
Country of study	USA
Age of learners	5-10: Third and fourth grade children (USA)
Type of study	**Researcher-manipulated evaluation: RCT (cluster)** Three experimental conditions, but no control.
Aims of study	'The purpose of the present study was to examine how to structure dialect instruction so that it is effective in teaching SE forms to students who use BEV in their writing.' (p 215)
Summary of study design, including details of sample	• N = 89 • Random allocation to 1 of 3 treatment groups • Pre-test (translation task and self-efficacy rating) • Teaching • Post-test (translation, comprehension, storytelling, self-efficacy) • Filtering (only students exhibiting BEV features in the translation pre-test were retained) • Analysis
Data-collection instruments, including details of checks on reliability and validity	*The translation task and the self-efficacy ratings were administered in Session 1 of the two training sessions. The testing lasted about15 minutes. In Session 2, the students completed a second self-efficacy measure, three post-test measures, then a third self-efficacy measure. The testing took about 35 minutes.* *The post-tests involved:* • *a translation task with different sentences (15 minutes)* • *a translation task with the same sentences as the pre-test (15 minutes)* • *a comprehension task (an oral story plus three questions)(8 minutes)* • *a story-writing task (12 minutes)* *The sessions were separated by one week.* Details of reliability: • The researcher trained teachers in the relevant forms of non-SE. The researcher was present during all training and testing of participants. Details of validity: • The ethnicity element was not mentioned and an informal check was made to establish whether the students were aware that the 'errors' presented to them were features of BEV. • The BEV grammatical features of the oral story and the translation tasks were established by reference to the linguistic research data. • The researcher ran teacher-training sessions to establish the instructions given to the students. • The researcher was present at all training and testing sessions.
	• It is stated (p 219) that the students understood the phrase 'correct grammatical form' as applying to written language (although no

	support evidence is given).
	• The teachers confirmed the researcher's view that the non-standard forms they had selected for training were problematic for many of their pupils.
Methods used to analyse data, including details of checks on reliability and validity	• Verbal argument plus ANOVA. • Analysis of variance (one-way and two-way ANOVA) and chi-square. • A Shapiro-Wilk Test of Normality was applied to the free-writing tasks before the ANOVA. More specifically: • Mean performance and standard deviation on pre-tests and post-tests were calculated as a function of training condition. Same on post-tested data, only a function of treatment condition. • Performance was calculated as mean percentage correct of SE forms. • Tukey pairwise comparisons with classes not students as the unit of observation. With self-efficacy ratings: • Two-way ANOVA was conducted with Treatment and Time of Test. • Means of self-efficacy ratings taken. • Chi-square analysis on disparity between treatment groups on pre/post test change in ratings. • ANOVAs with student as the independent variable. Details of reliability: • The pre-post translation test scores were analysed by classroom as well as by student. Details of validity: • The low pre-test means (3.48, 3.69 and 4.17 out of 12 max.) showed that the students had only translated a few BEV features into SE (p 221). • The reanalysis in terms of translation of the six grammatical categories indicated that improvement applied to five of them (p 223); thus the learning was not limited to one somehow salient feature of the input. • ANOVA is a fairly robust procedure, but given the smallish sample, the use of a normality test was a sensible precaution. • The free-writing tests were used to establish who reached a pass grade (65%) in a realistic task type. • The effect of Grade (3rd vs 4th) was explored with the self-efficacy ratings. p was >0.05 in all cases.
Summary of results	• H1 was supported. The students who had exposure, explanations and practice performed better. Moreover, 81% of the ESP students reached the acceptability criterion in free writing, versus just 55% of the ES students and 33% of the E students. • H2 was not supported. The students in the ESP condition tended to lower their efficacy ratings.
Conclusions	• Practice plus corrective feedback is necessary for effective learning of dialect grammar rules. • Short-term exposure is not enough to cause change at third or fourth grade. • Simply noticing a problem, or having it brought to your attention, can be counter-productive, unless accompanied by practice and feedback. • The ESP treatment also impacted on free writing which was not taught, but which is the ultimate criterion. • The decline in ESP self-efficacy ratings may be because the ESP students developed a sense of greater realism; realistic self-awareness may be an important step in learning - showing receptivity to taking remedial action and learning (p 229). • The initially high level of the self-efficacy scores may be an artefact, in that the students were responding globally and not to the six

	grammatical features tested (p 229). Globally, there is considerable similarity between BEV and SE.
	• The success of the ESP method may have been due in part to the focus on writing rather than oral skills; writing (a) allowed time for review and (b) may be more amenable to change.
	• The classroom techniques used can be adopted easily into existing syllabuses; they were 'clearly defined and easy to administer'.
	• The study was limited to a single and brief application of the method and the BEV features were not completely eliminated from SE text. Further research is required to explore the impact of longer and more comprehensive interventions (p 230).
Weight of evidence A (trustworthiness in relation to study questions)	**High:** The question was whether any intervention could improve students' writing (since no previous published method had worked well) and the study showed that ESP could. The only provisos were the small sample size (or specifically small cell size) and the use of a reduced number of BEV features.
Weight of evidence B (appropriateness of research design and analysis)	**High:** The design is appropriate for assessing short-term intervention of a small number of features.
Weight of evidence C (relevance of focus of study to review)	**High:** The question of getting students to alter features of non-standard oral dialect when writing formal prose is central to the review.
Weight of evidence D (overall weight of evidence)	**High**

Hilfman T (1970) Can second grade children write more complex sentences? *Elementary English* **47**: 209-214	
Country of study	Chicago, USA
Age of learners	5-10: First and second grade children
Type of study	**Researcher-manipulated evaluation: Pre- and post-test**
Aims of study	To help second grade children expand and elaborate their sentences by increasing their understanding of syntactic units, hoping that after exposure to word-form classes, through the use of sentence expansion techniques, they would write longer and more complex sentences (in that order) then before exposure.
Summary of study design, including details of sample	• N =19 (18 second grade and one first grade) • Pre- and post-test design with mid-test during intervention period
Data-collection instruments, including details of checks on reliability and validity	• *Titles of compositions given.* • *No other details about pre- and post-test* • *Some details given about mid-test* • *No details of reliability or validity*

Methods used to analyse data, including details of checks on reliability and validity	• For each pupil, the subordinate clause index for each of the three compositions was measured. After an S-C-I was recorded for each subject, the average clause length was determined. • % increase or decrease in S-C-I between pre- and post-test. • Reliability and validity addressed by reference to Hunt's determining of the S-C-I.
Summary of results	Figure 1 gives all the individual S-C-Is and average clause length in the pre-, mid- and post-tests. The results of the three compositions were compared for possible growth in complexity and length of sentence structure. Three graphs were constructed revealing that 10 pupils increased S-C-I, six pupils decreased and three remained the same. Average clause length improvement is reported but not commented on. • 10 increased S-C-I from January 31 to March 21 (52.63%) • 6 decreased S-C-I from January 31 to March 21 (31.58%) • 3 remained the same (15.79%) • The S-C-I seemed to increase as the average clause length increased and conversely.
Conclusions	The author concludes that 'the study provided some evidence that some second grade children can, through instruction, expand and elaborate their sentences by increasing their understanding of syntactic units. At the second grade level, more subjects were successful in writing longer sentences than in writing subordinate clauses. However, although evidence is not conclusive, the data seemed to show a slight increase in S-C-I, which appears normal when recognising the difficulty of obtaining a full page of writing at this age level.' (p 211).
Weight of evidence A (trustworthiness in relation to study questions)	**Medium to low**
Weight of evidence B (appropriateness of research design and analysis)	**Medium to low:** No control group so no controlling for extraneous variables. No controlling for regression to the mean effects or temporal effects.
Weight of evidence C (relevance of focus of study to review)	**Medium to low:** Conceptual focus, sample and measures all relevant but very brief details and examples given of intervention and compositions. Few details re context.
Weight of evidence D (overall weight of evidence)	**Medium to low**

McNeill JH (1994) Instruction for deaf students in syntactic cohesion. *Acehi Journal* 20: 88-95.	
Country of study	Assumed USA
Age of learners	11-16
Type of study	**Researcher-manipulated evaluation: pre- and post-test**
Aims of study	The researcher has aimed to find out whether 'natural language instruction '(p 89) of certain cohesion devices would help improve the 'syntactic cohesion' of the writing produced by four hearing-impaired students.
Summary of study design, including details of sample	• A small-scale pre-and post-test study of four individual deaf students in response to a single intervention designed to improve their writing skills. • The study employed a specially designed instructional intervention with a group of high-ability, hearing-impaired, high-school aged students aimed at increasing these students mastery of certain cohesive devices, thereby enhancing the clarity of their writing. • Measures to ascertain students' percentage rate of correct use of these devices were applied immediately before the intervention, immediately afterwards and after six weeks elapsed at the end of the intervention period.
Data-collection instruments, including details of checks on reliability and validity	• *10 papers were collected to represent students' work prior to the intervention.* • *Immediately after the intervention, the four assignments for the last week of the intervention period were used.* • *Six weeks after the intervention, the researcher asked to use two assignments written for another teacher for the purpose of analysis.* • No details of reliability or validity; the researcher appears to believe that the papers collected are representative of the work of each student at the particular point in time, and assumes that the papers collected will provide a valid indication of the ability of these students to use the cohesive devices targeted.
Methods used to analyse data, including details of checks on reliability and validity	• The researcher aimed to establish the percentage of correctly used cohesive devices. These were determined 'by calculating how often each of the five devices was used by each student, and then determining if each use was contextually correct or incorrect" (p 93). • No details of reliability. Two English teachers help perform the analysis before the intervention, but inter-rater comparability, for example, is not discussed. For later analyses, the researcher alone does the analysis. • Validity of the analysis is held to be self-evident.
Summary of results	• The average gain in correct use of target conjunctions by the students was from 60% pre-instruction to 87% post-instruction, and recorded at 83% six weeks after post-instruction. • Content analyses of the last 12 assignments for the history course showed that students' papers included longer phrases and sentences; there was more frequent use of 'however' and 'therefore'. • Maintenance data were higher than pre-instruction data, showing that students' learning was maintained. The author reported 'an increase in the percentage of correctly used devices as determined by comparing written assignments collected from the students before and after instruction' (p 93). Content analyses showed that immediately at the end of the intervention period 'students' papers included longer phrases and sentences, with sentences usually joined or set apart correctly by targeted terms' (p 94). Maintenance data 'showed that the students continued to use the five terms correctly more often than prior to instruction, yet not as high as those immediately following instruction' (p 94).

Conclusions	The author concludes (tentatively) that 'instructing deaf students with advanced written English skills in the use of cohesive devices in a non-English class setting' is successful (p 94). She views her research as replicable and suggests further avenues of research with deaf students.
Weight of evidence A (trustworthiness in relation to study questions)	**Medium to low:** The evidence is relatively clear, from this study. This is despite the small size of the sample and possible sources of bias.
Weight of evidence B (appropriateness of research design and analysis)	**Medium to low:** There is insufficient contextual detail to be sure the results are sound, and no detailed description of the sampling frame.
Weight of evidence C (relevance of focus of study to review)	**High to medium:** Again, highly relevant in principle but with insufficient scale, sampling context or explicitness to be generalisable.
Weight of evidence D (overall weight of evidence)	**Medium to low:** Basically a good small-scale study that should have led to a series of refined questions for further research, rather than to generalisations.

Roberts CM, Boggase BA (1992) Non-intrusive grammar in writing. Paper presented to the Annual Conference on Computers and Writing. Indianapolis, USA: May 1-3.	
Country of study	Presumed USA
Age of learners	11-16: Grade 10 (USA)
Type of study	**Researcher-manipulated evaluation: pre- and post-test**
Aims of study	The aim of the study is not particularly clear. One might deduce that the aim was to study whether the use of 'non-intrusive grammar instruction at the computer' would enhance students' ability to 'identify incomplete or unclear sentence structures' and their ability to identify 'sentence boundaries' (from abstract). The authors state the broad aims are: • for students to enjoy writing at the computer • for students to be able to write without initial concern for usage and spelling • to develop an awareness of the need for standard language usage • to concentrate on sentence boundary errors
Summary of study design, including details of sample	The study develops a particular intervention for an 'average' grade 10 class, based on a pedagogy of non-intrusive grammar instruction aimed at enhancing students' ability to identify and mark correctly sentence boundaries. Pre-intervention, during-intervention and post-intervention measures are used to measure the success of the intervention. N = 15 students
Data-collection instruments, including details of checks on reliability and validity	• *Assumed work collected by the teacher* • *No details of reliability or validity*
Methods used to analyse data,	• Identification of length and number of sentence-boundary errors.

The effect of grammar teaching (syntax) in English on 5 to 16 year olds' accuracy and quality in written composition

including details of checks on reliability and validity	• Simple word counts and counts of sentence-boundary errors. However, the study reports numbers of sentence-boundary errors only for the last piece of work analysed. • No details of reliability or validity.
Summary of results	The authors report significant gains in fluency and a reduction in sentence-boundary errors for a number of the students in the sample. Students became more fluent, measured in terms of word count and word gain. They did not all avoid sentence boundary errors, although 12 students of the 15 'appear to be checking and then revising their sentences'.
Conclusions	The authors conclude that their findings are significant, especially the finding that '12 students appear to be checking and then revising their sentences'. They conclude that their 'results' are 'so significant (and promising)...that the collaborative experiment will continue'; and that 'voice' is heard in the writing.
Weight of evidence A (trustworthiness in relation to study questions)	**Low**
Weight of evidence B (appropriateness of research design and analysis)	**Low:** The research design is inappropriate for gauging effectiveness. There is no pre-test or secure baseline. The intervention is poorly described.
Weight of evidence C (relevance of focus of study to review)	**Medium:** The non-intrusive approach to improving writing accuracy and quality is worthy of inclusion.
Weight of evidence D (overall weight of evidence)	**Low:** • The sample is too small. • The study is poorly conceived and vague. • Results are conflated with and confused with conclusions. • The aims are objectives are not clearly delineated, and do not lead to research questions or hypotheses. • The conduct of the study is ill-disciplined and the level of analysis is low.

Rousseau MK, Poulson CL (1985) Using sentence combining to teach the use of adjectives in writing to severely behaviorally disordered students. Unpublished research report: City University of New York.	
Country of study	USA
Age of learners	5-10 and 11-16
Type of study	**Researcher-manipulated evaluation: pre- and post-test**
Aims of study	To improve the quality of descriptive writing of behaviorally disoriented students.
Summary of study design, including details of sample	• N = 3 • Multiple-baseline across subjects • The treatments were sequential (Baseline - Treatment 1 - Treatment 2 - Treatment 3) • Descriptive praise and points were given at all stages • All stages had a sentence-combining part and a story writing part • The treatments differed with respect to (a) whether the sentence-combining periods focused on the same topics as the story-writing sessions, and (b) the focus of the praise and points: punctuation, adjectives, or different adjectives
Data-collection instruments, including details of checks on reliability and validity	• *Number of words per T-Unit* • *Number of adjectives per T-Unit* • *Number of different words per T-Unit* • *Comments on writing quality of stories* Details of reliability: • The researcher used a 17-point checklist. • Two raters were used. Details of validity: Not reported, beyond current work on sentence completion, and a common sense approach to increasing the physical rewards given for cooperating with the study.
Methods used to analyse data, including details of checks on reliability and validity	• Plotting 'adjectives per T-unit' and 'different adjectives T-unit' against sessions and fitting a line through the plots of the baseline and each of the three treatments. • Story quality was assessed by two graduate students - see instructions quoted on p 9. The person doing the assessment was asked to compare two pieces of writing (by the same student?): one written at baseline and one in a treatment phase (see page 14) • The data were graphed with a logarithmic scale to allow for the range of scores. • Presumably a simple least-squares procedure was used to obtain the line of best fit. Details of reliability: Inter-observer reliability was checked, using two raters, for a sample of 25% and found to be adequately high on average for • number of adjectives per T-unit (mean= 96%) • number of different adjectives per T-unit (mean= 97%) • T-unit length across all conditions (mean= 93%)

	• story quality (mean= 94% for 18 story pairs) • the procedures checklist (mean= 100%) Details of validity: The story raters were asked to indicate what they thought they were rating for and to comment on the quality of the stories. This acted as a validity check.
Summary of results	• There was a marked increase in the number of adjectives per T-unit when Treatment 1 was introduced. The results were maintained during Treatments 2 and 3 (from 0.16, 0.36 and 0.29 to 1.14, 2.58 and 1.13 at Treatment 1 for Chad, Andy and Joe) except one child, 'Chad' did not receive Treatment 2. • There was a marked increase in the number of different adjectives per T-unit with Treatment 1 and number continued to rise through Treatments 2 and 3 (from 0.15, 0.31 and 0.25 to an average of 1.04, 2.87 and 1.26 for the treatment sessions). • The mean number of words per T-unit increased by 2.56 and 3.34 words across the study as a whole for Chad and Andy. This represented four and five grade levels. Chad thus went from three grade levels below to one grade level above the norm, and Andy from one grade level below to four grade levels above. Joe's grade level did not change' (p 13). • The stories written during treatment(s) were judged to be better and (by one evaluator) more coherent as stories (more background information - taught - and more sequencing of actions - not part of the teaching).
Conclusions	• 'Improvement in the composition skills of academically deficient students was demonstrated as a function of reinforcement and • simple instructions' (p 14). This implies that the addition of sentence-combining instruction did not have an impact. • Students learned (or employed) rhetorical skills that were not being taught. In Treatment 1, adjective use was praised but not practised. In none of the treatment sessions was sequencing taught.
Weight of evidence A (trustworthiness in relation to study questions)	**Medium:** The problem is that there is not a clear research question. The lack of discussion about validating the tests also reduces the trustworthiness slightly.
Weight of evidence B (appropriateness of research design and analysis)	**Medium to low:** The multiple baseline design was reasonably appropriate, but the selection of the sample places it at medium to low.
Weight of evidence C (relevance of focus of study to review)	**Medium:** Three factors lower this to medium: the lack of detail about validating some of the instruments used, the lack of formulated research questions and the fact that the feedback appears not to have been linguistic (thereby inevitably emphasising the import/saliency of praise at the expense of language).
Weight of evidence D (overall weight of evidence)	**Medium to low**

Satterfield J and Powers A (1996) Write on! Journals open to success. *Perspectives in Education and Deafness* 15: 2-5.	
Country of study	Assumed USA
Age of learners	5-10 and 11-16

Type of study	**Researcher-manipulated evaluation** The intervention was a programme of teaching grammar that ran from October 1994 to March 1995.
Aims of study	To see whether a combination of whole language and traditional approaches could improve deaf students' acquisition of English as evidenced through their writing
Summary of study design, including details of sample	Study took five deaf students and, as part of their routine class teaching, delivered an intervention over six months. The intervention aimed to improve their written grammar by combining teaching of traditional grammar with whole language approach based on utilising pupils' own experiences.
Data-collection instruments, including details of checks on reliability and validity	*No details given*
Methods used to analyse data, including details of checks on reliability and validity	• Samples of raw data presented - no analysis made • No details of reliability or validity
Summary of results	• Students were encouraged to write more than previously. • Students retained grammatical concepts, 'making them a permanent part of their writing repertoire'. • 'Whole language approaches encourage students to concentrate of the process of writing, allowing them to transfer their thoughts to paper without focusing on every possibility for error.' • 'Whole language strategies, combined with mini-lessons in grammatical structure, enabled each of the five students to show improvement in written expression by the end of the school year.'
Conclusions	• It works: all five subjects' written expression improved by the end of the school year. • 'Deaf and hard of hearing students can indeed benefit from a learning environment that includes whole language principles' (p 5).
Weight of evidence A (trustworthiness in relation to study questions)	**Low:** Study has speculative interest only. Insufficient data, context and explicitness about analysis.
Weight of evidence B (appropriateness of research design and analysis)	**Low:** Poor research design and lack of any data analysis make this a very unhelpful study.
Weight of evidence C (relevance of focus of study to review)	**Medium:** Arguably the focus (improving written grammar of deaf children) is relevant, as is the concept of combining two approaches to teaching grammar.
Weight of evidence D (overall weight of evidence)	**Low:** Low level of execution in particular makes this a very lightweight study. The sample is small and unclear, there is a lack of methodological account, and the results and conclusions are fused.

Stock R (1980) The effect of teaching sentence patterns on the written sentence structures of grade two children. Unpublished report. Canada: Manitoba.	
Country of study	Reviewer infers USA.
Age of learners	5-10: second grade
Type of study	**Researcher-manipulated evaluation: controlled trial** Allocation was systematic. No mention of randomisation.
Aims of study	'The purpose of this study was to ascertain if the teaching and practicing of sentence building, by means of teaching specific capitalized and punctuated noun-verb sentence patterns, would result in the increase of these sentence patterns in the written narrative compositions of a sample of second grade children.' p 2
Summary of study design, including details of sample	Prospective allocation to experimental and control groups using matched pairs. The experimental group was taught a total of 15 lessons over a period of six weeks. The sentence pattern used was the noun-verb (subject-predicate) pattern. The lessons dealt with nouns, verbs, articles, adjectives, adverbs and how these elements combine into various noun-verb patterns. At all times, initial capitalisation and end punctuation were stressed as elements of a complete sentence. N = 19 matched pairs (i.e. 38)
Methods used to collect data	Seventy-five word samples of narrative writing.
Data-collection instruments, including details of checks on reliability and validity	No details given
Methods used to analyse data, including details of checks on reliability and validity	• Post-test T-unit ratio scores were calculated for each of the matched pairs. Also the mean and standard deviation of experimental and control groups. Mean number of T-unit scores for students in experimental and control groups. • No details of reliability or validity.
Summary of results	• The Null-hypothesis was not rejected: that is, the distribution of sign test values for each of the matched pairs did not differ significantly from equality. • Almost twice as many experimental students as control students increased their T1 units. The experimental group increased its use of the T1 unit from 20% to 52% compared with a percentage increase of 1.54% for the control group.
Conclusions	• Teaching and having children practice a specific noun-verb sentence pattern, complete with initial capitalisation and end punctuation did not result in a statistically significant increase in the use of the same pattern in children's writing. • The experiment did not result in a significant difference between the control group and the experimental group at the 0.05 level of significance.
Weight of evidence A (trustworthiness in relation to study questions)	**Low:** Insufficient information given about sample, measures, context, data-collection and analysis.
Weight of evidence B (appropriateness of research design and analysis)	**Medium:** Reviewer infers method of allocation is not random, therefore design is controlled trial and is appropriate to addressing question of effectiveness although not as appropriate as RCT.
Weight of evidence C (relevance	**Medium to low:**

of focus of study to review)	Lack of detail on conceptual focus, sample, context and measures.
Weight of evidence D (overall weight of evidence)	**Medium to low**

Stone AK, Serwatka TS (1982) Reducing syntactic errors in written responses of a retarded adolescent through oral patterning. *Education and Training in Mental Retardation and Developmental Disabilities* **17: 71-74.**	
Country of study	Assumed USA, as both authors are from Florida.
Age of learners	11-16: Although the single subject was a 14-year-old girl, the chronological age may be less relevant for a child defined as 'retarded'. However, the learning syntax may imply this age group or above.
Type of study	**Researcher-manipulated evaluation** A single-subject study of ABACA design
Aims of study	'To demonstrate a self-patterning, self-correction paradigm for transfer from spoken to written expression.' (p71)
Summary of study design, including details of sample	Single subject, ABACA design.
Data-collection instruments, including details of checks on reliability and validity	• No details of data-collection methods. • Reliability was checked during each phase by having another teacher randomly check responses against stated criteria. • Authors do not comment upon the validity of their data-collection methods.
Methods used to analyse data, including details of checks on reliability and validity	• Appear to have compared % change between five-day means of % errors. • Random checking of reliability by another teacher of experimenter's analyses using stated criteria. • No details of validity.
Summary of results	Following the ABACA design: • A showed a baseline of mean response of 50 words with 26% errors. • B showed a mean response of 52 words with 12% errors. • A showed a mean response of 53 words with 15% errors. • C showed a mean response of 60 words with 6% errors. • A showed a mean response of 62 words with 1% errors.
Conclusions	Authors conclude that 'syntactic errors could be reduced through a process of oral self-patterning with the teacher serving as a guide' (p74). (Reviewers' note: It was assumed that the subject's past failure to produce adequate written text was due to the lack of properly trained transfer skills from oral to written form.)
Weight of evidence A (trustworthiness in relation to study questions)	**Low:** The study is interesting, but really is so low in trustworthiness that many similar studies on far greater scales and with much more specific criteria, etc., are needed before anything may be taken from it as at all convincing.

Weight of evidence B (appropriateness of research design and analysis)	**Low**
Weight of evidence C (relevance of focus of study to review)	**Medium:** For this one subject, there seems to be some evidence that the research produced a desirable effect. However, only one subject was used, with no control students, and, with no information given about sampling to select this one subject, the generalisibility of this research is extremely low.
Weight of evidence D (overall weight of evidence)	**Low**

APPENDIX 4.2: Summary of weights of evidence for studies included in the in-depth review

Study	Weight of evidence A	Weight of evidence B	Weight of evidence C	Weight of evidence D
Bateman and Zidonis (1966)	Medium	High	Medium	Medium to high
Elley *et al.* (1975, 1979)	High	High to medium	Medium	High to medium
Fogel and Ehri (2000)	High	High	High	High
Hilfman (1970)	Medium to low	Medium to low	Medium to low	Medium to low
McNeill (1994)	Medium to low	Medium to low	High to medium	Medium to low
Roberts and Boggasse (1992)	Low	Low	Medium	Low
Rousseau and Poulson (1985)	Medium	Medium to low	Medium	Medium to low
Satterfield and Powers (1996)	Low	Low	Medium	Low
Stock (1980)	Low	Medium	Medium to low	Medium to low
Stone and Serwatka (1982)	Low	Low	Medium	Low

The weights of evidence A to D are defined in the EPPI-Centre guidelines on data-extraction (EPPI-Centre, 2002c) as relating to the following:

A: Trustworthiness in relation to study questions
B: Appropriateness of research design and analysis
C: Relevance of focus of study to review
D: Overall weight of evidence