Teaching argumentative non-fiction writing to 7–14 year olds

A systematic review of the evidence of successful practice

Review conducted by the English Review Group

Technical report written by Richard Andrews, Carole Torgerson, Graham Low, Nick McGuinn, Alison Robinson

EPPI CENTRE

SSRU Social Science Research Unit

INSTITUTE OF EDUCATION UNIVERSITY OF LONDON

Teaching argumentative non-fiction writing to 7–14 year olds

A systematic review of the evidence of successful practice

TECHNICAL REPORT

Review conducted by the English Review Group

Report by Richard Andrews (University of York)
Carole Torgerson (University of York)
Graham Low (University of York)
Nick McGuinn (University of York)
Alison Robinson (University of York)

The results of this systematic review are available in four formats. See over page for details.

The results of this systematic review are available in four formats:

SUMMARY   Explains the purpose of the review and the main messages from the research evidence

REPORT   Describes the background and the findings of the review(s) but without full technical details of the methods used

TECHNICAL REPORT   **Includes the background, main findings, and full technical details of the review**

DATABASES   Access to codings describing each research study included in the review

These can be downloaded or accessed at **http://eppi.ioe.ac.uk/reel/**

CONTENTS

Glossary

APU	Assessment of Performance Unit
CT	Controlled trial
DfES	Department for Education and Skills
EAL	English as an additional language
ICT	Information and communication technology
KS	Key Stage
NLS	National Literacy Strategy
PGCE	Postgraduate Certificate of Education
SRSD	Self-regulated strategy development
UKLA	United Kingdom Literacy Association
WoE	Weight of evidence

Abstract

Who wants to know?

Teachers, pupils and policy-makers are keen to know how best to teach and learn argumentative writing at Key Stages 2 and 3 in the National Curriculum. Difficulties with such writing are well-known.

What did we find?

Results showed that certain conditions are either assumed or have to be in place to create a climate for successful practice. These are not specific to argumentative writing but include:

- a writing process model in which students are encouraged to plan, draft, edit and revise their writing;

- self-motivation;

- some degree of cognitive reasoning training in addition to the natural cognitive development that takes place with maturation;

- peer collaboration, thus modelling a dialogue that (it is hoped) will become internal and constitute 'thought';

- and explicit and very clear explanations for students of the processes to be learned.

The specific strategies have been identified that have contributed to successful practice in teaching and learning with regard to argumentative writing for 7–14 year olds include:

- 'heuristics', i.e. scaffolding of structures and devices that aid the composition of argumentative writing – in particular, planning, which can include examining a question, brainstorming, organizing and sequencing ideas and evaluating;

- planning which is extensive, elaborated and hierarchical can make for more effective argumentative drafting and completion of essays;

- the use of oral argument, counterargument and rebuttal to inform written argument;

- the identification of explicit goals (including audiences) for writing;

- teacher modelling of argumentative writing;

- and 'procedural facilitation', i.e. coaching by the teaching through the process of writing argument.

How did we get these results?

Via an EPPI-Centre systematic review conducted in collaboration with policy-makers and teacher educators. The review question was: 'What is the evidence for successful practice in teaching and learning with regard to non-fiction writing (specifically argumentative writing) for 7-14 year olds?'

What are the implications?

The results from this review suggest that more work needs to be undertaken at key stage 2, in particular, to develop argumentative writing by linking it to critical thinking and other approaches which aim to encourage reasoning. While argument has a more secure place in the National Curriculum than in 1990, it needs to continue to be encouraged and developed; and more research needs to be undertaken to determine how best to teach it.

CHAPTER ONE
Background

1.1 Aims and rationale for current review

Our review focuses on the argumentative genres of non-fiction writing.

Non-fiction writing – as indicated by the negative definition – has been the least favoured aspect of writing in the English curriculum for many years. The first and second versions of the National Curriculum for England made little difference to this position, but the current version of the National Curriculum (DfEE, 1999) sets out a more balanced framework for writing and reading in which non-fiction takes its due place.

The reason for such neglect for much of the 20th century is that literature (especially fictional writing, such as the novel) formed the 'central civilizing presence' in the English curriculum. Most English teachers, at primary or secondary level, still see a literary core to their practice, values and professional training. The connection between personal development, the nurturing of the imagination, and the study of literature is still very strong in the minds of English teachers; indeed, as research in the 1990s showed (Goodwyn, 1992), the personal growth/literary model of English is the dominant one for English teachers. Such a tradition is a great and influential one, and we would not wish it to be diluted in any way. However, the question remains: what place does *non-fiction* hold in the curriculum?

1.2 Definitional and conceptual issues

It should be said that 'non-fiction' is an unduly negative term, defined in relation to fiction. Under the meta-genre of non-fiction sits a wide range of documentary and other genres or text-types: the essay, the report, the manual, the travel book, the travel guide and brochure, reportage, diaries, etc. For the purposes of the present study, for conven-ience's sake, we have continued to use the term 'non-fiction' to cover this range.

Non-fiction writing includes writing to inform, explain and describe (reports, explanations, manu-als, prospectuses); writing to persuade, argue and advise (essays, reviews, opinion pieces, advertise-ments); as well as writing to analyse, review and comment (commentaries, articles, etc.).

The present review, as justified later, focuses on the second two of these categories, excluding writing to inform, explain and describe. We character-ise the second two categories as broadly concerned with 'argumentative' writing.

1.3 Policy and practice background

To its credit, the English curriculum for the first part of the 21st century is fairly enlightened with respect to non-fiction. It is now no longer a prob-lem that non-fiction is absent from the English curriculum in all but the most formal and dry text-types. Rather, the latest version of the National Curriculum for English embraces a range of non-fiction forms alongside, and blended with, literary and expressive forms. The questions are now: what is the evidence for *successful* practice in the teach-ing of non-fiction, and how can we help teachers and learners to write non-fiction more successfully?

Although our focus is on *writing* non-fiction, and although reading and writing are framed separately within the National Curriculum for English, we take it as given that reading and writing are reciprocal activities, particularly in writing development. We also think that speaking and listening bear upon the writing of non-fiction, in that, for example, spoken forms of argumentation may well be better employed than they are now to help improvement in writing non-fiction.

We have focused on writing because (a) compe-tence in writing lags behind that in reading in Key

Stage (KS) 2 assessment tests, (b) less research has been done on writing than on reading and (c) pragmatically, in the timescale for the present review, we needed to focus on one particular aspect of the English curriculum.

At KS 2, for example, under Reading, pupils should be taught 'an understanding and appreciation of non-fiction and non-literary texts' (1999, p 54) and the various types of language that are embodied in them. Unfortunately, such types of texts are characterised as 'non-chronological' – a misleading and inaccurate term (see Andrews and Gibson, 1993) as many fictional works are non-chronological and many non-fictional ones chronological. The range of reading should include:

a. diaries, autobiographies, biographies, letters

b. print and ICT-based reference and information materials

c. newspapers, magazines, articles, leaflets, brochures, advertisements (p 55)

Similarly, for Writing, pupils should work in a range of forms including '…reports, explanations, opinions, instructions, reviews, commentaries' (ibid., p 58).

At key stages 3 and 4, the programme becomes more diversified and more specific. The range of reading is to include:

a. literary non-fiction

b. print and ICT-based information and reference texts

c. media and moving image texts

with examples of such texts being by authors such as Peter Ackroyd, James Baldwin, Flora Thompson, Dorothy Wordsworth, Alistair Cooke and Charles Darwin. Many of these are characterised as *literary* non-fiction, as if the curriculum is unwilling to let go of the literary dimension. In Writing at key stages 3 and 4, a range of purposes is set out, with forms such as 'memos, minutes, accounts, information leaflets, prospectuses, plans, records and summaries' included, as well as 'brochures, advertisements, editorials, articles and letters conveying opinions, campaign literature, polemical essays' (p 54).

It is important to note that the *functions* of writing at these secondary school key stages include persuading, arguing and advising, influencing the reader, analysing and reviewing, evaluating and presenting a case, as well as the more descriptive informing, explaining and describing. The distinction between 'argumentation' on the one hand, and 'description' on the other is an important one for our study, reflecting a high level but often simplistic categorisation between imaginative, descriptive and argumentative writing which derives from 19th century rhetorical theory and

which has influenced the writing curriculum ever since.

The introduction of the National Literacy Strategy into the primary sector in 1998 signalled a more decisive shift away from the orthodoxies of the 'cultural heritage' and 'personal growth' modes of English (Cox, 1991) towards a curricular model which foregrounded the explicit study of 'transactional' language (Britton, 1972) across and beyond the school curriculum. Drawing upon the work of Kress (1994) and others, the Strategy asserts the importance of young people being taught how to engage with the non-fictional genres they are most likely to encounter during their lives as citizens and workers.

Even as early as year 1, pupils are, therefore, introduced to such non-fictional texts as signs, labels, captions, lists and instructions. By the end of primary school, the range of non-fictional genres has broadened to include complex explanatory texts drawn from across the curriculum, as well as information and electronic texts (DfEE, 1998). This work is consolidated and developed at KS 3, where pupils are not only encouraged to explore the linguistic features and structures of a wide range of print, image and 'multi-modal' non-fictional texts, but also to shape their own creative engagement with these genres according to a threaded sequence of writing triplets which runs on into KS 4: 'imagine, explore, entertain … inform, explain, describe … persuade, argue, advise … analyse, review, comment' (DfEE, 2001a). Under the terms of the Strategy, the task of helping young people gain an understanding of how these non-fictional genres operate is perceived as being a responsibility for all teachers, not just those concerned with 'English'.

National strategies at primary level

In its Framework for Teaching Literacy (1998), the National Literacy Strategy (NLS) identified both termly fiction and non-fiction text ranges for all year groups from year1 to year 6 (a yearly overview for Reception classes) and text, sentence and word level objectives for both reading and writing. The initial launch of the NLS Framework was supported by national training programmes and resources to support effective teaching of both reading and writing. Resources in the first set of guidance and support materials for schools included a specific section on the teaching of non-fiction writing. Lewis and Wray were involved in writing the non-fiction objectives and the professional development materials to support the teaching (Lewis and Wray, 2000), which was at that time novel for the majority of teachers in England. The discursive text type was placed in the final term of year 4 and year 5 after children had two years' experience of writing 'non-chronological' reports and explanatory texts.

Additional materials to support the teaching of writing include *Developing Early Writing*, a hand-

book for practitioners in Foundation Stage and Key Stage 1 (DfES, 2001) and *Grammar for Writing*, a handbook and self-study CD Rom (DfES, 2000) (also supported by local authority consultant-led training programmes). In addition to these resources, web-based support material - such as the set of writing fliers designed to support effective teaching of both narrative and non-fiction writing – also promoted talk for writing and interactive teaching strategies to engage children in speaking and listening, and collaborative writing and drama as further ways to support writing development.

Planning exemplification units promote a teaching sequence from reading to writing, and supporting children in developing their own writing, having explored models of effective writing. Since the NLS became part of the wider Primary National Strategy, there has been a continued focus on supporting the teaching of writing. The publication of recent research undertaken with United Kingdom Literacy Association (UKLA) and several local authorities on approaches to improving boys' writing through the use of ICT is the most significant recent work on the teaching of writing (DfES, 2000).

National strategies at secondary level

The Secondary National Strategy for school improvement began life in 2001 as the Key Stage 3 National Strategy when the English and mathematics strands were introduced to all schools in England. Also at that time, the Strategy addressed literacy across the curriculum for teachers from all subject areas. From its onset, considerable guidance and support have been provided for teachers in secondary schools to improve pupils' writing. *The Framework for Teaching English: Years 7, 8 and 9* (DfEE, 2001a) contained teaching objectives for writing, including specifically those for writing to 'persuade, argue and advise'. *English Department Training* (DfEE, 2001b) and *Literacy Across the Curriculum* (DfEE 2001c) both had a section on 'Writing non-fiction' which considered a range of text types including 'persuasion'. *Literacy and Learning* (DfES, 2004a) followed up the Literacy across the curriculum resource in providing teachers of all subjects with guidance on using the teaching objectives from the Framework to assist them in developing pupils' literacy. *Improving Writing* (DfES, 2003, 2004b) was a major resource for English teachers which focused on aspects such as designing writing, text structure and organisation. The research that underpinned much of this work, particularly Improving writing, was the work of Bereiter and Scardamalia (1982, 1987), Derewianka (1990), Hillocks (1986, 1995), and Lewis and Wray (2000).

1.4 Research background

There have been concerns about the status of non-fiction writing in the English curriculum in England since the seminal study by Barnes et al. (1969), *Language, The Learner and the School* (1969). Although primarily focused on talk, that study – along with the seminal study by Moffett (1968) in the US, *Teaching the Universe of Discourse* - set the tone for a generation of research. It drew attention to the need for a balanced writing curriculum, leaning neither towards the dry, empty rhetorical genres that had become staple in the classroom in the 1950s and early 1960s, nor to the freer, more personal and 'creative' forms that had emerged in the 1960s. In the mid-1970s, two key research studies were published: *A Language for Life* (DES, 1975) and *The Development of Writing Abilities* (Britton et al., 1975), confirming the need for balance.

Research into argumentative writing took its lead within this context from Freedman and Pringle in Canada (1984), and Dixon and Stratta in England (1986). To focus on the development of thinking and practice in England, Dixon and Stratta trace their research back to 1979 when they began to study non-fiction and argumentative writing produced by young people for coursework examination for the then Certificate of Secondary Education (CSE). Working with the Southern Regional Examining Board, they discovered that 12-18 year olds were able to produce non-fiction work of high quality and imagination, and with the inclusion of a 'personal voice'. Significantly, they were reacting against a predominance of narrative in the curriculum. Their book, summing up six years of development, was entitled *Writing Narrative and Beyond*.

At the same time, from 1979 to the mid-1980s, the Assessment of Performance Unit (APU) had undertaken the largest ever survey of writing in England and Wales by 11 and 15 year olds, and come to the same conclusion: that non-fiction writing was under-represented in the curriculum and that, in particular, 11 and 15 year olds were not very good at argumentative writing in relation to their abilities with other modes of writing. The dearth of opportunity for imaginative writing of these kinds, the dominance of narrative, the reliance on conventional forms such as the essay, and the assumption that non-fiction writing was 'difficult' (on account of the conceptual load) manifested itself in the first version of the National Curriculum in English (1989) and in its modest revision (1995).

Against this background, Andrews began a PhD in 1987, completing it in 1992, on narrative and argumentative writing at year 8 (12-13 year olds) in three secondary schools. The results were partly negative, showing that it was not possible to build on narrative structural powers and understanding to write argumentatively; but positive, too, in that new forms of argumentative writing were tried successfully in the classroom. Pupils' understanding of the process of argumentative writing, their drawing on dialogic skills (and on speech genres) to compose argumentative writing, and the imaginative dimension of such writing were behind an action research project conducted in ten primary and

ten secondary schools in 1991-92, *Improving the Quality of Argument, 7-16* (Andrews and Costello, 1992), followed by a full report with evaluation (Andrews et al., 1993) which covered all the compulsory school years of 5-16. These and other pre-school and post-16 projects were collected in *Teaching and Learning Argument* (Andrews, 1995). The critical evaluation of these projects revealed a wide range of argumentative written forms that were accessible to 7-14 year olds (the focus of the present review).

During the 1980s, running alongside the gradual emergence of argumentation with narrative and other more personal forms of writing, there was increasing understanding of the writing process itself. Such understanding is best represented in the work of North American and Canadian researchers, such as Graves (1982) in his promotion and examination of documentary drafting and re-drafting by 7-11, year olds, and by Bereiter and Scardamalia (1987) in their development of writing process models. The basic principle of the pedagogic models that were developed was that, by understanding the writing processes of accomplished (literary and non-literary) writers, processes and procedures could be established for novice writers. It was understood that, whereas narrative writing was often accretive, non-fiction writing was more truly compositional (i.e. a question of 'putting things together' or com-posing) and thus suitable for planning and drafting. Word-processing packages are conducive to such kinds of composition due to the facility of moving around large chunks of text.

1.5 Authors, funders and other users of the review

The authors and Review Group consist of Richard Andrews and Carole Torgerson (co-co-ordinators of the English Review Group at York), Graham Low, Nick McGuinn and Alison Robinson. All have an interest in either the substance of the review (Andrews, Low, McGuinn, Torgerson), or the methodological approach of systematic reviewing (Torgerson, Robinson), or both. McGuinn has a working interest in teacher education, and is actively involved in the training of English teachers.

The advantage of four core reviewers (Andrews, Low, McGuinn and Torgerson) was that pairs could be formed to moderate decisions on keywords and extracted papers. Robinson, as Information Officer, conducted the literature searches, managed the review database, tabulated data for the review map, and co-ordinated the structure and content of the review report.

The Review Group was based at York in order to ensure efficiency and speed in the completion of the review.

The project was funded by the DfES through the EPPI-Centre, which is concerned with supporting reviews of research literature and encouraging their applicability. It is hoped that the results of this review will inform beginning and continuing teachers more fully about an important part of their subject.

The principal users of the review, in the course of its being undertaken, were teacher educators (McGuinn and Nicola Onyett who also works on the programme) and PGCE students in English at the University of York.

The main audiences for the completed review are likely to be teacher educators, trainee teachers and in-practice teachers. The review will also be of interest to teachers interested in research, policy-makers, researchers and pupils/students.

1.6 Review question

The core research question for the present review is:

> *What is the evidence for successful practice in teaching and learning with regard to non-fiction writing (specifically argumentative writing) for 7-14 year olds?*

and subsidiary questions include:

> *How does the evidence vary, if at all, by gender?*
>
> *How does the evidence vary, if at all, for pupils with EAL?*
>
> *Is there evidence to show ways in which pupils who have difficulty with this aspect of the curriculum can be helped to accelerate their progress?*
>
> *What aspects of teaching and learning best help pupils to improve their motivation for, and the quality of, their non-fiction (argument) writing?*
>
> *Is there evidence of barriers to and facilitators for progression and continuity in the development of non-fiction (argument) writing abilities across the transition from primary to secondary schools?*

The rationale for a focus on argumentative writing is partly cognitive, partly curricular and partly pragmatic. Cognitively, argumentative writing remains 'difficult' because it includes the operation and application of ideas – whereas one could argue that narrative or 'descriptive' writing deals with particularities. It is thus an important area in which to review research in an attempt to shed light on how best to help young people to think and to articulate that thinking more clearly. Such thinking and articulation are important across the curriculum, and within a democracy. In curriculum terms, although argumentative writing (and

non-fiction writing more generally) are far better represented in Curriculum 2000 than in the 1990 or 1995 versions of the National Curriculum, there remains uncertainty and lack of confidence among teachers as to how best to teach it (partly because most of them are trained in the literary tradition). Thirdly, and pragmatically, covering the continent of non-fiction writing in a one-year systematic review would require a large team and considerable resources. Our decision to begin by working on the argumentative aspects of non-fiction writing has provided us with a manageable project, but one which will also lay the foundations for further work in the non-fiction field.

The research focuses primarily at the whole text level, but takes into account research at sentence level, if relevant. It looks at research published internationally (between the years 1990 and 2005), but has as its immediate context the teaching and learning of argumentative non-fiction writing in England. It takes into account the frameworks provided by the National Curriculum for England, the National Literacy Strategy (KS 2) and the Framework for Teaching English Years 7 to 9 (KS 3). Where relevant, research evidence from KS 1 is included, although the main focus is at KS 2 and KS 3.

An implication of the research is: 'What do teachers need to know in terms of subject knowledge and subject application knowledge with regard to argumentative non-fiction writing?'

There are also implications for policy and for future research. The development of such writing skills for pupils is essential in helping to develop a critical voice, and for participation in the world of work and in democracies.

CHAPTER TWO
Methods used in the review

2.1 User involvement

2.1.1 Approach and rationale

In the summer term of 2005, we held a session with PGCE students in English at the University of York to review and develop the protocol. We introduced them to the systematic review methodology, but were principally interested in their views as beginning teachers on the focus and nature of the review.

Their principal contribution was to suggest that, in the wake of reforms to the National Curriculum for English at the end of the 1990s, the scope and range for non-fiction teaching within the curriculum had been increased and was now in balance with fictional, expressive and other more traditional forms of writing and reading in English lessons at key stages 3 and 4. Not only was there more balance: there was also an understanding among them that hybrid forms of text (e.g. novels that included non-fiction passages, advertisements using fiction, the use of narrative and story for non-fictional functions) were part of the creativity of reading and writing non-fiction. They reacted strongly to being 'boxed' by the curriculum, and found that in its present forms, the National Curriculum tended to compartmentalise genres and text-types.

The trainees also pointed out that, in their experience, some English departments were very wedded to the notion of a literary-based English curriculum, and 'taught non-fiction through fiction'. This was an issue we bore in mind as we investigated the most successful approaches to the teaching of non-fiction.

With the 2005-06 cohort that started in September 2005, we presented work in progress to the group as a whole and asked a small focus group from English PGCE to look at the emerging review in more detail, comment on its progress, and pro-

vide guidance as to its development. We also maintained contact with the 2004-05 cohort as they entered their first teaching jobs across the country, both by sending them the draft report for comment, and by inviting them back to York for discussion of its findings. In particular, we were interested in the implications for teaching.

2.2 Identifying and describing studies

2.2.1 Defining relevant studies: inclusion and exclusion criteria

The EPPI-Centre tools and guidelines for undertaking systematic reviews were used throughout the conduct of the review, in order to limit bias at all stages (EPPI-Centre, 2002a, 2002b and 2002c).

The review question looked for evidence of successful practice in teaching and learning with regard to argumentative non-fiction writing for 7-14 year olds. Therefore the relevant literature included studies that could be used to draw causal inferences: that is, inferences that various practices (strategies and methods) in the teaching and learning of argumentative non-fiction writing can improve pupils' non-fiction writing. Case studies, explorations of relationships and other non-experimental designs were included only where there was an evaluation.

The scope of the review was limited to English as a first, second or additional language; to students in KS 1-4; to articles or reports written in the English language; and to those published or unpublished but in the public domain between the years 1990 and 1995.

The full inclusion and exclusion criteria 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: ASSIA (Applied Social Sciences Index and Abstracts), BEI (British Education Index), C2SPECTR, ERIC (Educational Resources Information Center), PsycINFO, SSCI (Social Sciences Citation Index)

- Searching of reference lists of systematic reviews

Keywords for searching included the following:

- Argumentation, persuasive discourse

- Non-fiction, non-narrative, non-literary, discursive

- Expository writing, persuasive writing

Searches of these sources were limited so as to identify studies conducted in the time period 1990 to 2005.

The full search strategy for the electronic databases is contained in Appendix 2.2.

2.2.3 Screening studies: applying inclusion and exclusion criteria

The Review Group set up a database system, using EndNote, for keeping track of and coding studies found during the review. Titles and abstracts were imported and entered manually into the database. Pre-established inclusion and exclusion criteria were applied to (a) titles and abstracts, and (b) full reports. Full reports were obtained for those studies that appeared to meet the criteria or where we had insufficient information to be sure. The inclusion and exclusion criteria were reapplied to the full reports and those that did not meet these initial criteria were excluded.

2.2.4 Characterising included studies

The studies remaining after application of the inclusion/exclusion criteria were keyworded using the EPPI-Centre core keywording strategy (EPPI-Centre, 2002a). Additional keywords which were specific to the present review were added. All the keyworded studies were added to the larger EPPI-Centre database, REEL, for others to access via the website.

The EPPI-Centre and review-specific keywords are contained in appendix 2.3.

2.2.5 Identifying and describing studies: quality-assurance process

Internal QA: All four members of the Review Group were involved in a screening pilot where the inclusion and exclusion criteria were independently applied to a sample of 20 studies. Reviewers then met and discussed all the decisions and any implications for the process of double-screening the full database. Subsequently, application of the inclusion and exclusion criteria to all titles and abstracts at the first stage of screening and to all full papers at the second stage of screening was undertaken by pairs of reviewers, working first independently and then comparing their decisions and coming to a consensus. The keywording of studies included in the map was also undertaken in the same way.

External QA: At the first stage of screening, the EPPI-Centre link person (KD) independently screened a random sample of 40 titles and abstracts. We compared KD's decisions with the Review Group's moderated decisions. At the second stage of screening, the EPPI-Centre link person independently screened a random sample of nine papers. We compared KD's decisions with the moderated Review Group decisions. Similarly, at the keywording stage, KD double keyworded four papers with three members of the Review Group.

2.3 In-depth review

2.3.1 Moving from mapping to in-depth review

After completion of the keywording process, the Review Group met two members of the Advisory Group to reflect on the mapping of the field, and to decide whether any further inclusion/exclusion criteria might be applied. It was decided to retain the original question for in-depth review, but to narrow down to the highest quality evidence provided by the studies included in the map. Thus it was decided to include only randomised controlled trials and controlled trials in the in-depth review. In order to derive causal inferences, it is necessary to compare what happens when participants are involved in an intervention with participants in a control group not receiving the intervention, or receiving an alternative intervention. Studies that use a pre- and post-test design (i.e. without a control group) may be confounded by temporal effects or regression to the mean effects, which are controlled for in randomised controlled trials and partially controlled for in controlled trials.

2.3.2 Detailed description of studies in the in-depth review

The in-depth review reported and synthesised findings from the included randomised controlled trials and controlled trials.

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

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

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 were based on:

A   Soundness of studies (internal methodological coherence), based upon the study only

B   Appropriateness of the research design and analysis used for answering the review question

C   Relevance of the study topic focus (from the sample, measures, scenario, or other indicators of the focus of the study) to the review question

D   An overall weight taking into account A, B and C

2.3.4 Synthesis of evidence

The data were synthesised to bring together the studies which answered the review questions and which met the quality criteria relating to appropriateness and methodology.

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 were conducted by pairs of Review Group members working first independently and then comparing their decisions and coming to a consensus. The EPPI-Centre link-person also contributed to quality assurance by double data extracting one study with each member of the Review Group (total of four studies).

CHAPTER THREE
Identifying and describing studies: results

3.1 Studies included from searching and screening

This systematic review included study types B (exploration of relationships), C (evaluation) and E (review) as defined in the EPPI taxonomy of study types (EPPI-Centre, 2002a). The term 'study' is taken to mean one complete piece of work. The term 'paper' is taken to mean the research report or article in which the study is reported. Single studies may sometimes be reported in more than one research paper. The initial review stages of searching and screening focused on research papers. Where studies identified for inclusion in the map were found to be reported in more than one paper, the study was included and keyworded using data from all of the papers in which it was reported. The map is therefore presented as an overview of characteristics of included studies.

Table 3.1 gives the origin of all papers found and those subsequently included in the systematic map.

Papers found on ERIC, PsycINFO, SSCI, BEI, ASSIA and C2SPECTR by application of the search strategy described in section 2.2.2 were imported and de-duplicated sequentially into the review database in the order shown in Table 3.1. One additional potentially relevant paper was identified through expert contact. Titles and abstracts were screened in accordance with the inclusion and exclusion criteria described in section 2.2.3. Potentially included papers were sent for and then screened again on the basis of the full paper.

The screening process identified 29 papers that met the inclusion criteria. Three studies were each found to be reported in three pairs of papers, resulting in a map of 26 included studies. Three of the studies were reviews and 23 were primary studies. The conclusions of the three reviews were summarised to help inform the map and synthesis (see Table 3.2), and the 23 primary studies were keyworded to establish their main characteristics (see tables 3.3 to 3.5).

Figure 3.1 illustrates the process of filtering papers from searching to mapping and finally to synthesis.

3.2 Characteristics of the included studies (systematic map)

3.2.1 Summary of conclusions of reviews

Three reviews met the inclusion criteria for the map, all of which were narrative reviews. No systematic reviews were identified.

Crowhurst (1990) aimed to examine evidence about students' performance in writing persuasive/argumentative discourse, consider questions of difficulty and development, and suggest appropriate teaching strategies. Gleason (1999) aimed to address the role of evidence in argumentative writing and to examine techniques for improving the soundness and strength of students' arguments, with a view to the sustainability and feasibility of the interventions. Newcomer and Barenbaum (1990) reviewed literature pertaining to the written composing skills, including expository writing skills, of students with learning disabilities. They present implications and suggestions for teaching and research in this area.

3.2.2 Main characteristics of the included primary studies

Although we searched and screened for explorations of relationships and other non-experimental designs, Figure 3.2 shows that all of the studies that met the inclusion criteria for the systematic map were researcher-manipulated evaluations. Of the 23 included studies, 16 were trials (7 controlled trials and 9 randomised controlled trials), and five were of a pre- and post-test design. The remaining two were correlational studies.

Figure 3.1 Filtering of papers from searching to map to synthesis

STAGE 1
Identification of potential studies

One-stage screening
papers identified in ways that allow immediate screening, e.g. handsearching

Two-stage screening
Papers identified where there is not immediate screening, e.g. electronic searching

1,630 citations identified

Title and abstract screening

Citations excluded
Criterion 1	1,270
Criterion 2	9
Criterion 3	32
Criterion 4	70
Criterion 5	0
Duplicates	136
TOTAL	**1,517**

1 citation identified

113 citations

114 citations identified in total

STAGE 2
Application of exclusion criteria

Acquisition of reports

9 reports not obtained

105 reports obtained

Full-document screening

Reports excluded
Criterion 1	47
Criterion 2	2
Criterion 3	14
Criterion 4	11
Criterion 5	2
TOTAL	**76**

26 studies in 29 reports included

STAGE 3
Characterisation

Systematic map
of 26 studies (in 29 reports)

Studies excluded from in-depth review
Not RCT or CT	7
Reviews	3
TOTAL	**10**

STAGE 4
Synthesis

In-depth review
of 16 studies

Tables 3.3, 3.4 and 3.5 categorise the 23 included primary studies by study type and describe their main characteristics.

Table 3.3 gives the main study characteristics of the nine randomised controlled trials, including details about the nature of intervention.

Eight of the nine studies were conducted in the US; one was carried out in Canada.

The majority of the RCTs (N=7) involved students in the equivalent KS 2 age range of 7-11 years and five involved students in the KS 3 age range of 11-14. Four studies focused solely on students at the equivalent KS 2 level and two studies solely on students at KS 3. Three studies involved students across both age groups.

Four of the studies involved solely students of mixed ability and three involved solely students with learning disabilities. Two focused on both mixed ability students and those with learning disabilities.

In five studies, the interventions lasted six weeks or less. One intervention was applied over five months and one over seven months. The length of intervention was not stated in two studies.

Seven of the studies reported pre- and post-test writing outcomes, and two reported post-test outcomes only.

The English Language context was not stated in five studies. In four studies, English was being taught to first language students. In one study, English was being taught as both a first and as a second/additional language.

Table 3.4 gives the main study characteristics of the seven controlled trials (CTs), including details about the nature of intervention.

Six of the seven studies were conducted in the US and one in Canada.

The CTs were slightly more biased towards the equivalent KS 3 age range than the RCTs. Four studies involved students at the equivalent KS 2 level and five involved students at KS 3. Two focused solely on students in the KS 2 age group and three solely on students on students at KS 3. Two studies involved students across both age groups.

Six studies involved students of mixed ability and one focused solely on gifted students.

Four of the seven studies involved interventions lasting six weeks or less. In one study, the intervention lasted for eight weeks. One study applied the intervention for one year. In one study, the length of intervention was not stated.

Six studies reported pre- and post-test writing outcomes and one reported post-test outcomes only. One study reported pre- and post test results, and formative and summative measures of writing outcomes.

In all seven CTs, English was being taught to first language students. In one study, English was being taught as both a first and as a second/additional language.

Table 3.5 gives brief study details of the remaining seven studies included in the map. These studies were of a pre- and post-test or correlational design. Such studies are less reliable than RCTs or CTs because their design does not control for temporal or regression to the mean effects, or for selection bias.

Five of the seven studies were conducted in the USA and two in Canada.

Two focused solely on the KS 2 age range and four solely at KS 3 level. One study involved students in both age groups. Three studies involved students of mixed ability and four focused on those with learning disabilities.

In three cases the interventions lasted three weeks or less. Two studies applied interventions over nine and 13 weeks respectively. One intervention was applied over nine months. The length of intervention was not stated in one study.

Six studies reported pre- and post-test writing outcomes and one reported post-test only. In six studies, English was being taught as a first language. The English language context was not stated in the remaining study.

3.3 Identifying and describing studies: quality-assurance results

Screening of titles and abstracts - internal QA: All titles and abstracts retrieved from all of the electronic searches were independently double screened by two reviewers who then met and resolved any disagreements.

Screening of titles and abstracts - external QA: The EPPI-Centre link person (Kelly Dickson) independently screened a random sample of 40 titles and abstracts. Kelly's decisions were compared with the Review Group's moderated decisions. We agreed on decisions on 33 papers and disagreed on decisions on seven papers. We calculated the Kappa statistic for inter-rater agreement, which was 0.44 (moderate). It was not necessary to take the process further because of the rigorous method of internal QA.
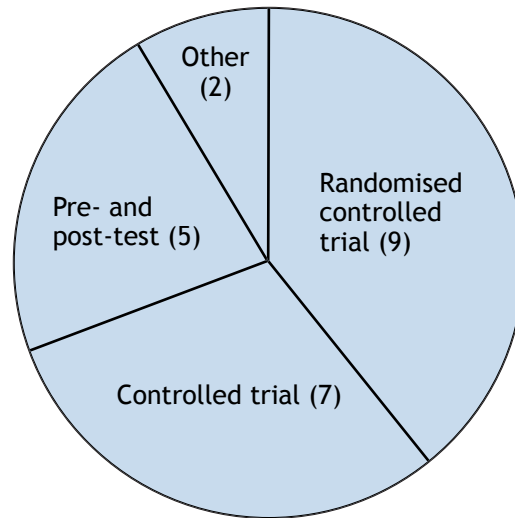
Screening of papers - internal QA: All papers sent for after first stage screening were screened by two members of the Review Group. Agreement was extremely high. For the few papers where

Table 3.1 Origin of included studies

	Found	Number of papers included	Number of studies on which included papers reported	Reviews	Primary studies
ERIC	611	18	15	3	12
PsycINFO	455	10	10	-	10
SSCI	263	1	1	-	1
BEI	131	-	-	-	-
ASSIA	20	-	-	-	-
C2SPECTR	14	-	-	-	-
Contact	1	-	-	-	-
Total	*1,495*	*29*	*26*	*3*	*23*

Table 3.2 Main conclusions of included reviews with regard to teaching and learning strategies

Author (date)	Main conclusions with regard to teaching and learning strategies
Crowhurst (1990)	'It is not enough to instruct students in the structures and linguistic forms of argument and to provide assignments for practice… Effective teacher intervention will consist of providing opportunities to practice contextually relevant writing, and giving instruction to students, individually or in groups, when they are ready to profit from it.' (p 357)
	The author further concludes:
	• Students should be encouraged to select issues they feel strongly about.
	• They should direct their persuasive writing to teachers, classmates, principals and others.
	• They should engage in group discussion of issues and do pre-writing to clarify their thoughts.
	• They should read, as well as write, persuasive/argumentative writing; students will more easily acquire linguistic forms and structures of argument if they are exposed to models of them.
	• Discussion of such readings should cover both content and structure.
Gleason (1999)	'A key to teaching argumentation ... is to provide explicit instruction about the role of evidence, how to develop a line of reasoning that supports a conclusion, and how to use factual information for that purpose. Students must also learn to search for the information that would assist them in developing their evidence. Another key is to provide feedback to students using criteria based on an established model of argumentation and emphasising coherence of argument.' (p 102)
Newcomer and Barenbaum (1990)	'Possibly the extent of the problem for writers with learning disabilities can be attributed to a dearth of instruction or practice in special education classrooms…it is equally plausible that students with learning disabilities receive instruction and opportunities to practice but lack the cognitive or linguistic capacity either to retain these organizational cues or to generalize their use from one task to another.' (p 587)
	They report that 'the evidence…presents a bleak picture of the learning disabled writer of expository text' but note that 'much of the research was conducted by the same group of investigators' (p 588).
	The authors conclude overall that 'it is essential that students be committed to the idea of improving writing and be capable of taking control of the writing process. Otherwise, regardless of the instructional program, revisions are superficial rather than substantive'. It remained for teachers 'to help their charges learn to care about writing and see a need to do it well' (p 591).

Fig 3.2 Type of researcher-manipulated evaluation (N = 23, mutually exclusive)

there was initial disagreement, full agreement was achieved through discussion.

Screening of papers - external QA: The EPPI-Centre link person (KD) independently screened a random sample of nine papers. Kelly's decisions were compared with the Review Group's moderated decisions. We agreed on eight and disagreed on one - this is very high agreement, but, in any case, we also double-screened every paper at the second stage to use the most rigorous approach for internal QA. It was therefore not necessary to take this process further.

Keywording - internal QA: Keywording of all papers included in the systematic map was undertaken by two reviewers, working independently and then meeting to resolve any disagreements.

Keywording - external QA: External quality assurance was provided by KD, the EPPI-Centre link person. Agreement on the four studies independently keyworded by KD and CT or RA or NM was good and any differences were resolved through discussion.

3.4 Summary of systematic map

Three reviews and 23 primary studies were included in the systematic map. The three reviews are summarised in Table 3.2. Of the 23 primary studies:

- All were researcher-manipulated evaluations

- Nine were randomised controlled trials, seven were controlled trials and seven were other types of study design.

- Nineteen studies were conducted in the USA and four in Canada.

- Eight studies involved students solely within the equivalent KS 2 age range and nine involved students solely within the equivalent KS 3 range. Six studies involved students across both age groups.

- Thirteen studies focused solely on students of mixed ability and seven studies focused solely on students with learning disabilities. One focused on gifted students. Two studies involved both mixed ability students and those with learning disabilities.

- Twelve studies involved interventions lasting six weeks or less. Three studies applied interventions over eight, nine and 13 weeks respectively, and three studies applied interventions over five, seven and nine months respectively. In one study, the intervention lasted for one year. The length of intervention was not stated in four studies.

- Nineteen studies reported pre- and post-test writing outcomes, including one study which reported pre- and post test results with formative and summative measures of writing outcome. Three studies reported post-test writing outcomes only.

- In 15 studies, English was being taught to first language students only. In two cases, it was being taught to both first and second/additional language students. In six studies, the English language teaching context was not stated.

Table 3.3 Characteristics of randomised controlled trials (RCTs) (N = 9)

Author, date, country	Age and ability	Length of intervention	Writing outcome	English Language context
Crowhurst (1990) Canada	11-12 (Grade 6) Mixed attainment	Twice a week for five weeks	Pre- and post-test results	Not stated
De La Paz and Graham (1997) USA	10-13 (Grades 5, 6 and 7) Learning disabilities	Not stated	Pre- and post-test results	As a first language
De La Paz and Graham (2002) USA	12-14 (Grades 7 and 8) Mixed attainment	Six weeks	Pre- and post-test results	As a first language
Englert et al. (1991) USA	9-11 (Grades 4 and 5) Mixed attainment Learning disabilities	Seven months	Pre- and post-test results	Not stated
Ferretti et al. (2000) USA	9-12 (Grades 4 and 6) Mixed attainment Learning disabilities	Not stated	Post-test results	Not stated
Graham et al. (2005) USA	8-9 (Grade 3) Learning disabilities	Five months	Pre- and post-test results	As a first language and as a second/ additional language
Knudson (1991) USA	9-13 (Grades 4, 6 and 8) Mixed attainment	Two weeks	Post-test results	Not stated
Knudson (1992; 1994) USA	8-11 (Grades 3 and 5) Mixed attainment	Two weeks	Pre- and post-test results	Not stated
Troia and Graham (2002) USA	9-11 (Grades 4 and 5) Learning disabilities	Approximately 10 hours	Pre- and post-test results	As a first language

Table 3.4 Characteristics of controlled trials (CTs) (N = 7)

Author, date, country	Age and ability	Length of intervention	Writing outcome	English Language context
Burkhalter (1994; 1995) USA	9-11 (Grades 4 and 6) Mixed attainment	Up to one month	Pre- and post-test results	As a first language
Hammann and Stevens (2003) USA	13-14 (Grade 8) Mixed attainment	Six days	Pre- and post-test results	As a first language
Hidi et al. (2002) Canada	Mainly 11-12 (Grade 6) Mixed attainment	Eight weeks	Pre- and post-test results	As a first language and as a second/ additional language
Reznitskaya et al. (2001) USA	9-11 (Grades 4 and 5) Mixed attainment	Five weeks	Post-test results	As a first language
VanTassel Baska et al. (1996) USA	9-12 (Grades 4, 5 and 6)	One year	Pre- and post-test results	As a first language
VanTassel-Baska et al. (2002) USA	7-15 (Grades 2 to 9) Gifted	Not stated	Pre- and post-test results; Formative and summative measures	As a first language
Yeh (1998) USA	12-13 (Grade 7) Mixed attainment	Six weeks	Pre- and post-test results	As a first language

Table 3.5 Characteristics of other types of researcher-manipulated evaluation (N = 7)

Author, date, country	Age and ability	Length of intervention	Writing outcome	English Language context
Pre- and post-test study design				
Aulls (2003) (Study 1 only) Canada	12-13 (Grade 7) Mixed attainment	Thirteen weeks	Pre- and post-test results	As a first language
De La Paz (1997) USA	10-11 (Grade 5) Learning disabilities	Three weeks	Pre- and post-test results	Not stated
De La Paz (2001) USA	13-14 (Grade 8) Learning disabilities	Not stated	Pre- and post-test results	As a first language
Hallenbeck (1999; 2002) USA	12-13 (Grade 7) Learning disabilities	Nine months	Pre- and post-test results	As a first language
Sexton et al. (1998) USA	10-12 (Grades 5 and 6) Learning disabilities	Six to eight hours	Pre- and post-test results	As a first language
Correlational study design				
Cox et al. (1991) USA	8-11 (Grades 3 and 5) Mixed attainment	Two sessions	Post-test results	As a first language
Gordon (1990) Canada	11-12 (Grade 6) Mixed attainment	Nine weeks	Pre- and post-test results	As a first language

CHAPTER FOUR
In-depth review: results

4.1 Selecting studies for the in-depth review

The systematic map yielded 23 studies that met the inclusion criteria contained in Appendix 2.1. In order to establish the highest quality evidence provided by the studies in the map, the inclusion and exclusion criteria were narrowed down according to study design, as described in section 2.3.1. This process identified sixteen studies for in-depth review. (See Table 4.1.)

Seven studies (Aulls, 2003; Cox et al., 1991; De La Paz, 1997; De La Paz, 2002; Gordon, 1990; Hallenbeck, 1999/2002; Sexton et al., 1998) were excluded from the in-depth review because they were not of a randomised controlled trial or controlled trial design. These studies could not be reliably used when addressing the research question as their design did not control for temporal or regression to the mean effects, or for selection bias.

Summary tables giving further details of the studies included in the in-depth review are contained in Appendix 4.1.

4.2 Assessment of weights of evidence

Eleven of the 16 studies identified for in-depth review were rated as medium quality or above in terms of overall weight of evidence: Englert et al. (1991) and Ferretti et al. (2000) were rated 'high'; De La Paz and Graham (1997) and Troia and Graham (2002) were rated 'high to medium'; De La Paz and Graham (2002) and Graham et al. (2005) were rated 'medium to high'; and Crowhurst (1990), Knudson (1991), Knudson (1992; 1994), Reznitskaya et al. (2001) and Yeh (1998) were rated 'medium'.

Five studies were rated of low to medium quality or below: Hamman and Stevens (2003), Hidi et al. (2002) and Van Tassel-Baska et al. (2002) were rated 'low to medium'; and Burkhalter (1994;

1995) and Van Tassel-Baska et al. (1996) were rated 'low'.

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

The 11 studies rated of medium quality or above form the basis of the synthesis. It was decided only to include studies judged by the Review Group to have an overall weight of evidence judgement (WoE D) of 'medium' or above on the basis that this quality of evidence could be relied upon in terms of the reliability and validity of the research. Studies with overall judgements of 'medium to low', 'low to medium' or 'low' were thought not to be able to provide particularly reliable or valid evidence. A study rated as 'high' would represent the highest quality of internal validity, be of a highly appropriate research design for our research question, and be highly relevant to the review in terms of the sample, context and measures. A study rated as 'medium' would be included, but caution would be urged in interpreting the results, as there are likely to be some limitations in the internal validity, the appropriateness of the research design, the relevance to our review, and the choice of sample, context and outcome measures. Similarly, studies in the intermediate categories between 'high' and 'medium' could have some shortcomings in one or more of the categories.

Table 4.3 shows the main characteristics of the studies included in the synthesis, ranked by their overall weights of evidence. All six studies rated 'medium to high' or above were randomised controlled trials. Three of the five studies rated 'medium' were randomised controlled trials and two were controlled trials.

4.4 Synthesis of evidence

The research question which the review attempts to answer is 'What is the evidence for successful practice in teaching and learning with regard to

Table 4.1 Studies identified for in-depth review (author, year, title)

Burkhalter N (1994; 1995) A Vygotsky-based curriculum for teaching persuasive writing in the elementary grades

Crowhurst M (1990) Reading/writing relationships: an intervention study

De La Paz S, Graham S (1997) Effects of dictation and advanced planning instruction on the composing of students with writing and learning problems

De La Paz S, Graham S (2002) Explicitly teaching strategies, skills, and knowledge: writing instruction in middle school

Englert C, Raphael T, Anderson L, Anthony H, Stevens D (1991) Making strategies and self-talk visible: writing instruction in regular and special education classrooms

Ferretti R, MacArthur C, Dowdy N (2000) The effects of an elaborated goal on the persuasive writing of students with learning disabilities and their normally achieving peers

Graham S, Harris K, Mason L (2005) Improving the writing performance, knowledge, and self-efficacy of struggling young writers: the effects of self-regulated strategy development

Hammann L, Stevens R (2003) Instructional approaches to improving students' writing of compare-contrast essays: an experimental study

Hidi S, Berndorff D, Ainley M (2002) Children's argument writing, interest and self-efficacy: an intervention study

Knudson R (1991) Effects of instructional strategies, grade, and sex on students' persuasive writing

Knudson R (1992; 1994) An analysis of persuasive discourse: learning how to take a stand

Reznitskaya A, Anderson R, McNurlen B, Nguyen-Jahiel K, Archodidou A, Kim S (2001) Influence of oral discussion on written argument

Troia G, Graham S (2002) The effectiveness of a highly explicit, teacher-directed strategy instruction routine: changing the writing performance of students with learning disabilities

VanTassel-Baska J, Johnson D, Hughes C, Boyce L (1996) A study of language arts curriculum effectiveness with gifted learners

VanTassel-Baska J, Zuo L, Avery L, Little C (2002) A curriculum study of gifted-student learning in the language arts

Yeh S (1998) Empowering education: teaching argumentative writing to cultural minority middle-school students

argumentative non-fiction writing for 7-14 year olds?' In the US and Canada, where all the studies in this in-depth review took place, 'argumentative non-fiction writing' is sometimes categorised as a sub-section of *expository* writing. It has been important, in the course of the review, to make sure that any expository writing that has been examined is indeed argumentative, and not mere exposition or description. A 'how to' or descriptive paper would be excluded; but a compare/contrast, opinion (supported by evidence), or persuasive paper would be included.

4.4.1 High rated studies

Both Englert et al. (1991) and Ferretti et al. (2000) were rated 'high' overall in terms of weight of evidence.

The study by Englert et al. (1991), 'Making strategies and self-talk visible: writing instruction in regular and special education classrooms' examines the effects of an intervention 'that attempted to improve students' expository writing abilities through an instructional emphasis on student dialogues about expository writing strategies, text structure processes, and self-regulated learning'

(p 337). The study was undertaken with 4th and 5th grade students, in the US. The intervention consisted of training in planning, organising, writing, editing and revising different text types. The writing process model is derived from a standard model that emerged in the 1980s in North America in the wake of work by Graves (1982), and Bereiter and Scardamalia (1987) on writing process: that is, that the taught and learnt model should reflect the writing processes of experienced writers. It is also based on a specific programme, the Cognitive Strategy Instruction in Writing (CSIW) which was 'designed to incorporate many features of effective strategy instruction, including the development of students' metacognitive knowledge about writing strategies through an emphasis on teacher modelling of an inner dialogue for directing the writing process, scaffolded assistance…procedural facilitation…through the use of think-sheets, and peer collaboration in writing conferences' (p 342).

The emphasis on text structures focuses attention not only on the shape and structure of a piece of writing, but also on making the implicit structures explicit to emergent writers. The results of the study showed that students who were exposed to the CSIW treatment showed increasing understand-

Table 4.2 Weights of evidence (WoEs) of studies identified for in-depth review

Author, date	WoE A	WoE B	WoE C	Overall weight of evidence (WoE D)
Burkhalter (1994, 1995)	Low to medium	Low	Medium	Low
Crowhurst (1990)	Medium	Medium	Medium	Medium
De La Paz and Graham (1997)	High	High to medium	High to medium	High to medium
De La Paz and Graham (2002)	Medium	Medium	High	Medium to high
Englert et al. (1991)	High	High	High	High
Ferretti et al. (2000)	High	High	High	High
Graham et al. (2005)	High	Medium	Medium	Medium to high
Hamman and Stevens (2003)	Low	Medium	Medium	Low to medium
Hidi et al. (2002)	Low	Medium	Medium to low	Low to medium
Knudson (1991)	Medium	Medium	Medium	Medium
Knudson (1992, 1994)	Medium	Medium	Medium to low	Medium
Reznitskaya et al. (2001)	Medium	Medium	Medium	Medium
Troia and Graham (2002)	High	High	Medium	High to medium
Van Tassel-Baska et al. (1996)	Low	Medium	Medium	Low
Van Tassel-Baska et al. (2002)	Low	Medium	Medium	Low to medium
Yeh (1998)	Medium to low	Medium to low	Medium to high	Medium

Table 4.3 Main characteristics and overall weights of evidence of studies included in the synthesis

Author, date, country	Study design	Age of participants	Overall weight of evidence (WoE D)
Englert et al. (1991) USA	RCT	9-11 (Grades 4 and 5)	High
Ferretti et al. (2000) USA	RCT	9-12 (Grades 4 and 6)	High
De La Paz and Graham (1997) USA	RCT	10-13 (Grades 5, 6 and 7)	High to medium
Troia and Graham (2002) USA	RCT	9-11 (Grades 4 and 5)	High to medium
De La Paz and Graham (2002) USA	RCT	12-14 (Grades 7 and 8)	Medium to high
Graham et al. (2005) USA	RCT	8-9 (Grade 3)	Medium to high
Crowhurst (1990) Canada	RCT	11-12 (Grade 6)	Medium
Knudson (1991) USA	RCT	9-13 (Grades 4, 6 and 8)	Medium
Knudson (1992, 1994) USA	RCT	8-11 (Grades 3 and 5)	Medium
Reznitskaya et al. (2001) USA	CT	9-11 (Grades 4 and 5)	Medium
Yeh (1998) USA	CT	12-13 (Grade 7)	Medium

ing and command of the structures underlying text, as well as a growing sensitivity to their audiences and to their purposes in writing. One of the findings, for example, showed that compare/contrast texts were significantly easier for students to organise than explanations, although the reverse was true in terms of their writing voice and sensitivity to the audience. The implication is that managing the 'voice' in argumentative writing is more difficult, and identifying the audience is also more difficult, perhaps because due to the relative formality of the task and the uncertainty over who is speaking/writing to whom.

One aspect of the results of this study was that students with learning difficulties performed neither better nor worse than those without such difficulties. This is an important finding in that, in the 1980s, students with learning difficulties were not often exposed to the complexity of a writing model, such as CSIW, it being assumed that they would progress more readily with programmed, limited and instructed procedures.

There are also suggestions in the paper that the skills learnt by the experimental group were transferable across the different types of writing undertaken: explanations, compare/contrast and 'expert writing'. The control group seemed not able to make such transfers across text-types.

The authors conclude that 'the data from the present study suggest that instruction in the writing process and expository text structures can be effective when they are embedded in an instructional framework emphasising teacher modelling, scaffolded assistance, procedural facilitation, peer collaboration, and the development of an inner language and vocabulary for talking about writing' (p 369).

Ferretti et al. (2000), in a more recent study undertaken in the US, aimed to investigate 'the effects of giving students an elaborated goal that included explicit sub-goals based on the elements of argumentation as compared with a general goal to convince an audience to agree with their opinion' (p 695). Specifically, 4th and 6th graders in the general goal groups were asked to write a letter to persuade an audience to agree with them on a position, whereas those in the experimental groups were asked to use the following explicit subgoals: a statement of their belief, two or three reasons for their belief, examples of supporting information, two or three reasons why others might disagree and why those reasons were wrong.

The 6th graders in the experimental group included more of the subgoals and strategies in their writing and thus wrote more persuasively than their control group counterparts. The 4th graders wrote equally persuasively in both conditions and included equal numbers of argumentative elements in both essays. Again, both students with and without learning difficulties appeared to

benefit from the more specific instruction. The difference between the performance of grade 6 students and those in grade 4 was not attributed by the researchers to developmental differences; one explanation put forward by the study is that the difference may be to do with the combined effects of composing and at the same time meeting the elaborated (more specific) subgoals; or the fact that 6th graders already have a more developed schema for oral and written argument which was reflected by the specific elements of argument that were used in the intervention.

The paper concludes that, overall, 'normally achieving students and those with [learning difficulties] may benefit from instruction on goal setting' (p 700) but the authors also suggest that the essays in themselves were not very persuasive, and only half (54%) of the sixth-grade students used rebuttals or alternative positions in their arguments. They suggest that 'the provision of explicit goals, along with intensive, scaffolded instruction in cognitive strategies and self-regulatory strategies…may help all students write more persuasively' (ibid).

Full details of interventions and summaries of results are contained in the summary tables in Appendix 4.1.

4.4.2 High to medium, and medium to high rated studies

Both De La Paz and Graham (1997), and Troia and Graham (2002) were rated 'high to medium'.

De La Paz and Graham's study aimed to examine the effects of dictation and explicit instruction in advanced planning on the writing of opinion essays by 5th, 6th and 7th grade students with learning difficulties. Students received instruction in either (a) planning, where they were taught strategies for developing, evaluating and organising ideas prior to composition, or (b) comparison, where students were taught *about* essay structure, revised sample essays, and composed and shared essays with fellow students. Half the students in each group composed their essays orally, while the other half wrote their plans and essays. The most effective combination for these students was that of dictation (oral composition) and instruction in advanced planning (rather than teaching *about* argumentative structures), reflected in the fact that these students wrote more complete and qualitatively better essays than those in the other groups and conditions. These results were measured in a post-test and two weeks later, in order to gauge the sustained effect (or not) of the intervention.

Two further aspects of the results are worth reporting: that those students taught the advanced planning techniques (as opposed to those who were taught *about* essay structure) spent more time in planning; and that whether the students dictated or wrote their compositions did not affect

Table 4.4 Details of interventions and summaries of results of high rated studies

Englert et al. (1991)

Nature of intervention

The intervention involved training in planning, organising, writing, editing and revising differ-ent text types. Work was individual and collaborative, and guided by a series of think-sheets. Assessment involved testing (1) metacognitive knowledge about the writing process, (2) skill at writing explanation and comparison essays, and (3) ability to transfer skills to reading and compre-hension of expository text (p 348).

The treatment group received systematic instruction in text analysis, modelling the writing proc-ess, guided practice and independent use of strategies (p 350). They wrote a class paper, an indi-vidual paper, then a paper for a class book (p 352).

The control group received regular writing instruction, which involved some planning, brainstorm-ing and collaboration, but was not systematic and the ideas were not made 'visible'. They wrote texts 2-3 times a week (p 352).

Summary of results

There were significant main effects for treatment, group and text, and a significant interaction between group and treatment, but no other significant interactions. The treatment effects seemed to be attributable to gains in students' holistic ratings, primary traits, and sensitivity to the read-ers. All these differences favoured the CSIW treatment, suggesting that students in the CSIW treat-ment showed increasing mastery of the structures underlying text, as well as growing sensitivity to their audience and purpose for writing.

Ferretti et al. (2000)

Nature of intervention

'Students in the general goal condition were asked to take a position and write a letter to per-suade an audience to agree with them.

'Students in the elaborated goal condition were given the same general goal plus explicit subgoals based on the elements of argumentative discourse.

'Subgoals directed students to include (a) a statement of their belief, (b) two or three reasons for their belief, (c) examples or supporting information for each reason, (d) two or three reasons why others might disagree, and (e) why those reasons were wrong'. (p 696)

Summary of results

Persuasiveness: There was a significant grade by goal effect, 'because the elaborated goal enhanced the writing of the sixth-grade students but not that of the fourth-grade students….' In addition, normally-achieving students wrote more persuasively than did students with LD, and papers about homework were more persuasive than those about violence' (p 697).

Elements of argument: There was a significant interaction between grade and goal, attributed to the same reason as for persuasiveness. There was also a significant interaction between grade and disability status, because students with LD included fewer elements than normally achieving stu-dents at Grade 4 (but not Grade 6).

the number of propositions they included in their essays. The authors are at pains to point out that dictation (oral composition) itself did not make for advances in composing skill, but that the combination of oral composition and advanced planning techniques made the difference. They also make the caveat that the study was conducted with students with learning difficulties, and may not necessarily be generalised to 'their normally achieving peers' (p 220).

The study, however, appears to suggest that direct use of heuristics or techniques for planning argumentative writing, combined with oral composition (thus freeing the students from the labour of writing their essays) was the most effective set of approaches. In this sense, there is some common ground with Englert et al. (1991) and Ferretti (2000), discussed above, both of which found that the use of explicit 'scaffolding' had an effect on students' argumentative writing.

Graham was also involved in a study of the effectiveness of a highly explicit, teacher-directed instructional routine used to teach three planning strategies for writing to 4th and 5th grade students with learning difficulties (Troia and Graham, 2002). The strategies used in this study included identifying the purposes of the activity and setting clear goals; brainstorming ideas; and organising those ideas. An acronym, STOP & LIST, was used to facilitate teaching of these elements: *s*top, *t*hink *of p*urposes, *l*ist *i*deas, *s*equence *t*hem. The writing process itself was divided into four stages: writing a rough draft, revising the draft, proofreading and editing, and publishing the final version. Teachers identified multiple tasks and situations for which the students could use the strategies, and gave students homework in which they could apply the strategies. Feedback was given on each completed assignment.

The authors found that there were no significant differences between groups in post-test scores for either essay quality or essay length. More specifically, the post-test essays written by students in the strategy instruction group were slightly longer but of lower quality than their pre-test essay, whereas the post-test essays written by the students in the writing process group (the control group) improved slightly in quality, but were shorter in length than the essays written for the pre-test. Two caveats must be borne in mind with this study: first, that the results are based on only three homework exercises, so it may be that the instruction hardly had time to have a significant effect on the learners; and, again, the fact that the study was undertaken with students with learning difficulties means that it may not be generalisable to a wider population of students of this age.

Unlike the previous two studies (but slightly lower in overall weight of evidence), the study by De La Paz and Graham (2002) was conducted with 12-14 year olds at grades 7 and 8, and covered the

full range of abilities. The aim in this case was to examine 'the effectiveness of an instructional program designed to improve the writing performance of (American) middle school students' (p 687). The key element of the instruction was 'a strategy that organized and directed the processes for planning and writing an essay' (ibid.). The strategy included developing a plan in advance of the writing that analysed the demands of the writing assignment; setting goals for writing; and generating and organising material to write about. The students also planned *while* they wrote, revising and upgrading their original plan as necessary, including transition words, interesting or mature vocabulary, and varied (error-free) sentence types.

As expected, the writing programme 'had a positive effect on the writing performance of the participating…students. Immediately following instruction, students in the experimental group produced essays that were longer, contained more mature vocabulary, and were qualitatively better than the essays generated…in the controlled classrooms' (pp 695-696) and these effects were maintained on an essay written a month after the instruction ended. The essential elements of the planning process, according to the authors, were that the 'plans of the students in the experimental condition tended to be more complete, elaborate and hierarchical' (p 696) than those in the control condition. Effect sizes were greater than 1.0 on both the post-test and maintenance writing probes (the tools used to test whether the effect was sustained).

Graham's work appears again in Graham et al. (2005), a study which aimed to examine 'the effectiveness of an instructional program designed to improve the performance of struggling young writers…attending urban schools that serve minority and other children from mostly low income families' (pp 208, 234). (The study assumes a connection between minority children and those from low income families.) Working within a self-regulated strategy development (SRSD) approach to learning, which emphasises that learning 'is a complex process that depends, in large part, on changes that occur in the learner's strategic knowledge, domain-specific knowledge, and motivation' (p 208), the students were taught strategies for accomplishing specific writing tasks, and any information or skills needed to use these strategies. There was thus a high degree of self-directed and teacher- and peer-supported development in this study. The specific planning strategy taught to these students was represented by the mnemonic POW: *p*ick my ideas, *o*rganise my notes, and *w*rite and say more. As part of the central organising stage with regard to persuasive essays, a second mnemonic, TREE, was used: *t*ell what you believe (i.e. state the proposition or 'topic sentence'), give three or more *re*asons (to support why you believe this), *e*xamine each reason (why will my reader buy it?), and *e*nd it (write a conclusion).

Table 4.5 Details of interventions and summaries of results of high to medium and medium to high rated studies

De La Paz and Graham (1997)

Nature of intervention

Students were taught a validated strategy for developing, evaluating and organising ideas for their essays in advance of composing. Students were encouraged to modify and supplement this initial plan (as needed) while composing. Students assigned to the comparison condition received no instruction in advanced planning, but rather they learned about the characteristics of good essays for meaning and structure, and composed and shared their own essays with peers.

A unique feature of the study involved the use of dictation as a means for generating not only the essay but also the students' advanced plans.

Summary of results

The combination of dictation and instruction in advanced planning resulted in more complete and qualitatively better essays in contrast to those written by students in the comparison condition on both a post-test and a two-week maintenance probe.

At post-test and maintenance, students who received advanced planning instruction spent more time planning than students in the two comparison conditions.

Students assigned to the two advanced planning conditions did not differ significantly in the number of propositions in their plans at post-test or at maintenance.

Propositions: Students assigned to the two AP conditions did not differ significantly in the number of propositions in their plans at post-test or maintenance (one-way ANOVA post-test $p=0.149$; maintenance: $p=0.491$).

Transformations: The two advanced planning conditions were not significantly different as post-test ($p=0.876$) or maintenance ($p=0.718$).

Essay length: Students in the AP condition who dictated composed longer essays than students in the CW condition.

Essay elements: 'essays composed by students in the planning condition who dictated contained more functional elements than students in the CW condition at post-test ... and at maintenance' (pp 214-215).

Essay coherence: 'students in the AP and dictation condition composed more coherent essays than students in the CW condition' at post-test.

Essay quality: AP students who dictated composed qualitatively better essays than CW students at post-test. AP students (writing or dictation) composed qualitatively better essays than CW students at maintenance.

Rate: CD students composed faster than AP writing students, or CW students (i.e. than students who wrote) at post-test. CD students composed faster than AP writing students.

AP students composed faster via dictation than writing at maintenance.

Troia and Graham (2002)

Nature of intervention

Instructors modelled how to use the three strategies (goal-setting, brainstorming, and organising) to perform several different types of tasks (including story-writing), explaining how the strategies were adapted for each particular task and how they affected performance.

Instructors identified multiple tasks and situations for which students could use the strategies.

Students were given homework assignments in which they applied the strategies to activities other than story-writing. For these assignments, instructors provided students with advice on how to apply the strategies to these new tasks and gave them feedback on each completed assignment.

Summary of results

For stories: Students who were taught to use the strategies wrote stories that were qualitatively better than those produced by their peers assigned to the process writing condition.

For essays: There were no significant differences between groups in post-test difference scores for essay quality or essay length. The post-test essays written by children in the strategy instruction group were slightly longer but of lower quality than their pre-test essay. In contrast, the post-test essays written by students in the process writing group improved slightly in overall quality, but were shorter in length compared with essays written prior to instruction.

Product: No significant differences between the groups for essay quality (p=0.18) or length (p=0.57). Post-test essays slightly longer and of lower quality.

Process: No significant group differences for essay planning time (p=0.79). 'Students in both groups (strategy and process) spent little or no time (less than one minute) on advance planning for their pre-test or post-test essays' (p 298).

Propositions were 0, as students wrote no initial plans.

De La Paz and Graham (2002)

Nature of intervention

The programme primarily focused on teaching students strategies for planning, drafting and revising text. The knowledge and skills needed to support these processes were also emphasised.

The key element of the instructional programme was a strategy that organised and directed the processes for planning and writing an essay. The students developed a plan in advance of writing that involved analysing the demands of the writing assignment, setting goals for writing, and generating and organising material to write about.

With SRSD, students are explicitly taught writing strategies along with procedures for regulating these strategies and the writing process. The procedures include goal setting, self-monitoring, and self-instruction.

Summary of results

In comparison with peers in the control condition, students in the experimental condition produced essays that were longer, contained more mature vocabulary, and were qualitatively better. At post-test, the plans of students in the experimental condition were better developed.

Planning: There was a significant main effect for trials (p<0.01) and instructional condition (p=0.00; effect size effect size: 1.17; maintenance effect size: 1.04).

Word length: There was a significant main effect for instructional condition (p<0.01; post-test effect size: 0.82; maintenance effect size: 1.07). Experimental students wrote longer essays.

Vocabulary: There was a significant main effect for trial (p=0.04) and instructional condition (p<0.01; post-test effect size: 1.13; maintenance effect size: 0.94). Experimental students used more long words.

Quality: There was a significant main effect for instructional condition (p=0.00; post-test effect size: 1.71; maintenance effect size: 0.74). The experimental students wrote tests of higher overall quality.

Graham et al. (2005)

Nature of intervention

With the SRSD model, the students were explicitly and systematically taught strategies for accomplishing specific writing tasks. Students were also taught any information or skills needed to use these strategies.

Students were also taught how to use self-regulation procedures, including goal- setting, self-monitoring, self-instructions, and self-reinforcement to help manage the target strategies and task of writing, as well as to obtain concrete and visible evidence of progress.

With the SRSD plus peer support, the concept was introduced whereby students would act as partners to help each other apply the strategies they were learning to other situations and in other classes.

In the comparison condition writing instruction was delivered to students by their regular teacher using the writers' workshop model.

Summary of results

SRSD instructed students wrote qualitatively better stories and persuasive papers (the two instructed genres) than their peers in the comparison condition. The effects of SRSD instruction generalised to one of the uninstructed genres, as students in both conditions wrote informative papers that were qualitatively better than their counterparts in the comparison condition.

Composing time: Students in the two experimental groups spent more time composing their post-test persuasive essays than students in the comparison condition.

Length: There was a statistically significant difference between the three conditions in the length of persuasive essays.

Elements: There was a statistically significant treatment effect. Students on the two SRSD conditions included more basis elements than their counterparts in the comparison condition but there was no statistical difference between the SRSD conditions.

Quality: There was a statistically significant treatment effect. Students in both SRSD conditions wrote qualitatively better persuasive essays than comparison students.

The results of this study demonstrate that students using the experimental SRSD-informed strategies wrote qualitatively better and longer essays than their peers in the comparison condition. The experimental students also spent more time composing their post-test essays. In general, the authors conclude that 'teaching third grade struggling writers a general strategy for planning a composition, genre-specific strategies for…persuasive writing, procedures for regulating these strategies and the writing process, as well as knowledge about the basic purpose and characteristics of the [genre] had a powerful effect on the participating writers' performance' (p 234). However, the study was not able to follow up the students' persuasive writing with a delayed post-test, so the authors were not able to claim that the significant effects of the intervention were sustained beyond the period of the experiment itself.

Full details of interventions and summaries of results are contained in the summary tables in Appendix 4.1.

4.4.3 Medium-rated studies

Crowhurst (1990), Knudson (1991), Knudson (1994), Reznitskaya et al. (2001) and Yeh (1998) were rated 'medium' in terms of weight of evidence. We have decided to include these in our synthesis as, in their various ways, they shed helpful light on the research question in hand, despite some shortcomings in methodological validity and/or reliability.

Crowhurst's study is one of the few undertaken in Canada in the present review. It aimed to discover whether students' 'writing of persuasion' (p 157) could be improved by instruction, and specifically whether practice in reading improves writing and vice-versa. 11-12 year old students (Canadian grade 6) were divided into four groups, each of which received a different combination of input. The first group underwent training in writing instruction, with the provision of a model structure, an opportunity for collaborative brainstorming, draft revision in pairs and teacher feedback on four 'for and against' essays. The second group had the same as the first group, plus the addition of reading five specially constructed 'for and against' texts. The third group read the same texts as the second group, then discussed them – but had no writing instruction. The fourth group acted as a control group, with discussion only and no extra input to the writing process.

The results indicate that the first two of the three experimental groups scored significantly higher than the control group on the writing quality at the post-test stage but not on the pre-test. Specifically, the post-test compositions of the first two groups 'were better organized, with fewer reasons – some of them elaborated – than the list-like compositions common in pre- and post-test compositions by students without instruction. Post-test compositions of the writing and reading+instruction groups were more likely to have some kind of concluding statement as against the very abrupt endings common

in other compositions' (pp 166-167). There were no significant differences between the groups on the number of idea units recalled.

Knudson (1991) worked with students in grades 4, 6 and 8 in southern California. There were three types of intervention used: instruction with model pieces of writing, scales and questions designed to guide students' writing and revision; both models and scales/questions; and no explicit instruction in persuasive writing (the control group). Results confirmed the difficulties of argumentative writing for students of this age, showed some improvement in content and form, and appeared to be moderately highly correlated with regard to clarity, coherence, organisation and word choice. Grade 8 students wrote better arguments (i.e. improved more) than those in grades 4 and 6, and were also able to sustain performance two weeks after the intervention. But this result in itself is not that surprising, and the author herself concludes that the results were mixed and inconclusive, and that there were limitations in study design. She also acknowledged that there was nothing to explain why girls' scores dropped so dramatically as soon as the intervention was withdrawn.

A later study by Knudson (1992, 1994) describes work with grade 3 and 5 students using a similar intervention to the previous study. This time, there were no significant main effects for gender, although there were significant main effects for grade. As in the study by Knudson mentioned above, such a result is not surprising and seems to point toward cognitive maturation being a significant factor in the ability to write persuasively, rather than any intervention on the part of the teacher. Knudson concludes that 'little is really known about what makes a good persuasive argument' and 'even less is known about how to teach effective argumentation' (p222). Unfortunately, in neither study is there a clear account of the interventions used.

The study by Reznitskaya et al. (2001) aimed 'to provide evidence about the effects of discussions in which children engage in oral argumentation on (sic) the reasoning that the children then exhibit in persuasive essays' (p 157). It examines 'whether oral discussions can help students acquire "portable" [i.e. transferable] knowledge of argumentation' (p 159). The intervention in this study consisted of discussion of controversial issues, coaching by teachers in formal argument devices, and web forums with grades 4 and 5 (9 to 11 year olds) – a series of interventions that went under the umbrella of 'collaborative reasoning'. At the end of the intervention period, students from the experimental and control groups each wrote a persuasive essay based on a moral dilemma. The essays were coded to measure students' ability to consider a variety of relevant arguments, counter-arguments and rebuttals, as well as to use evidence and to employ certain formal argument devices. Not surprisingly, students who had participated in collaborative reasoning discussions wrote essays

that contained a significantly greater number of arguments, counter-arguments, rebuttals, uses of formal argument devices, and references to text information [evidence] than the essays of similar students who did not experience the intervention. The results, however, must be treated with caution, as the authors themselves, acknowledge, as the study was quasi-experimental.

The final paper included in this synthesis is that by Yeh (1998). The aim of his study was to investigate the effectiveness of two heuristics (scaffolding devices) based on Toulmin's (1958) model of argument and on elements of classical rhetoric. The study was conducted with 12-13 year olds, specifically 'cultural minority' middle school students in two different schools in the San Francisco Bay area. The interventions were plans and scaffolds for writing argument, or 'devices to teach students a pattern of thought' (p 53). Their intention was to achieve a well-formed essay 'that avoids focus on superficial aspects of the written product' (ibid.). In order to focus on the deeper aspects of composition, the first heuristic devised was a 'pyramid', closely modelled on Toulmin, with a thesis, claim or proposition connected to, and supported by, data or evidence via an explicit or implicit 'warrant' (the part of the argument that justifies the connection between the evidence and the claim). The second heuristic was a 'bridge' linking the reason to an opinion via facts, if/then statements and values.

The results show that gain scores were higher in the experimental groups than in the control groups as far as argumentative development and 'voice' were concerned, but not significantly higher with regard to command of the conventions for argumentative writing. The gains were also higher for cultural minorities than for the majority of white students. From the questionnaire/survey results that accompanied the experimental element, it appeared that Hispanic- and African-American were less aware of the thesis-support model than White students (Asian-American students were excluded because of the small sample), although Yeh acknowledges that a more balanced sample of white and minority ethnic students would be needed to confirm these findings. Overall, the findings suggest that combining explicit instruction in heuristics with immersion (process) approaches to writing development are important, especially for minority ethnic groups.

In general, it appears that there are two aspects that need to be in place to ensure that improvement in argumentation takes place: first, the conditions must be in place to underpin the interventions (i.e. the cognitive programmes, writing practices and other factors that appeared to be necessary to the success of the interventions); and, second, the actual interventions must occur to improve argumentative writing *per se*.

All the studies above the 'medium' weight of evidence category suggested that specific heuristics

Table 4.6 Details of interventions and summaries of results of medium rated studies

Crowhurst (1990)

Nature of intervention

Students were divided into four groups to examine different combinations of input/skills:

Writing instruction

Instruction = provision of a model structure, collaborative brainstorming, draft revision in pairs, + teacher feedback of four 'pro' and 'con' essays.

Reading + Instruction

Students saw the writing structure model. Instruction emphasised the readings' structure and content. They read five specially-written paired pro-con texts and discussed them - inclduing identifying the structural elements.

Reading + Discussion

Instruction was in group discussion skills. They read the same texts as Group 2.

Control (discussion)

Instruction in group discussion. Students did group tasks.

Summary of results

Reading: No significant differences were found.

Writing quality:

1. Significant main effect for test ($p < 0.001$) and text x group interaction ($p < 0.01$). At pre-test, reading+instruction group scored below the control group. At post-test, the writing and the reading+instruction groups scored higher than the control group. This confirmed hypothesis 1a and 1b, but not 1c.

2. Significant main effect for judge; 'considered not to be significant since the effect was evenly distributed across groups and tests' (p161)

3. ANOVAs on specific features of the essays of the writing, reading+instruction and control groups showed: number of reasons/100 words: significant main effect for Test ($p < 0.001$). The number of reasons was generally smaller for the post-test than the pre-test; number of elaborations/100 words: significant main effect for test ($p < 0.05$) and Test x Group interaction ($p < 0.001$).

4. Writing group used fewer elaborations than the control group at pre-test, but more at post-test; number of conclusions: significant main effect for Test ($p < 0.001$) and Group ($p < .01$), plus significant Test x Group interaction ($p < .001$). No differences at pre-test, but at post-test, the writing group used more conclusions.

Knudson (1991)

Nature of intervention

There were four treatments, as described above:

1. Instruction with model pieces of writing

2. Scales and questions designed to guide students' writing and revision

3. *Both* model pieces of writing *and* scales and questions

4. The control group was shown a picture every day and asked to write a story about what was happening. They were not given explicit instruction in persuasive writing.

Summary of results

For both writing samples, four of the scores appear to be moderately highly inter-correlated: clarity, organisation, coherence and word choice.

There was a significant main effect for grade for the content score in that 'grade 8 wrote better than students in grades 6 and 4, and students in grade 6 wrote better than students in grade 4 for Writing sample 1. 'For Writing sample 2, however, students in grade 8 wrote significantly better than students in grades 6 and 4, but there were no significant differences between the mean scores of students in grades 6 and 4' (p 147).

For the form score, 'there was a significant main effect for grade' (ibid.).

Furthermore, 'although girls wrote significantly better than boys at the conclusion of treatment, after 2 weeks boys and girls received similar scores on their essays' (ibid.): that is, girls dropped off in performance two weeks after the treatment.

Knudson (1992, 1994)

Nature of intervention

The intervention consisted of four elements:

1. Presentation of model

2. Presentation of scales/questions/criteria

3. (1) and (2) combined

4. Free writing plus, oral interaction for some classes

Summary of results

In terms of classification of persuasive statements, analyses revealed that 'there were no significant main effects for gender', although there were 'significant main effects for grade'. There was 'no significant interaction effect for Gender X Grade'. The 'number of strategies employed increased significantly with grade' (p 220). There was a 'significant difference in the use of two categories for the dominant message by grade: Category 5 (Compromise) and Category 24 (Simple Statement)' (p 220).

'Analyses on the use of types resulted in similar findings. Type 3, Negative Sanction, was used more by students in Grade 5 ... than by students in Grade 3 ... Type 4, Request, was used significantly more by students in Grade 3 ... than by students in Grades 10 ... or 12' (p 220).

(Details are provided on pages 220 and 221.)

The use of reason 'was a frequently used dominant strategy'. There was 'a significant main effect for gender ... but no significant main effect for grade ... There was no significant interaction effect for Gender X Grade ... girls used significantly more reasons than boys' (p 221).

The results for the analyses of reasons used are depicted in Table 3 (p 222). 'The students appealed to the welfare or good of specific people or groups of people under the general categories of Safety, Convenience, or Pleasure when asking for a change in a school rule. Convenience was the most frequently used category, and Other Students/Friends were cited most often as the ones who would benefit from the change in a school rule' (p 221).

'There were no significant effects for treatment or for oral interaction for the instructional intervention' (p 221).

Reznitskaya et al. (2001)

Nature of intervention

The intervention consisted of discussion of controversial issues; coaching by teachers in formal argument devices; and web forums.

Students in the CR classrooms met twice weekly in small groups of between 6 and 8 participants to discuss controversial issues for between 15 and 20 minutes. Teachers coached these students in the specific formal argument devices that promote the development of reasoned. The CR students also engaged in 15-minute CR discussions with the other participating classrooms via Web forums. At the end of the five-week intervention, students from CR and contrast classrooms wrote a persuasive essay based on a moral dilemma. The essays were coded to measure students' ability to consider a variety of relevant arguments, counter-arguments and rebuttals as well as to use evidence and to employ certain formal argument devices.

Summary of results

'In this study, students who participated in CR discussions wrote essays that contained a significantly greater number of arguments, counterarguments, rebuttals, uses of formal argument devices, and references to text information than the essays of similar students who did not experience CR' (p 171). The essays written by CR students also contained more words.

Yeh (1998)

Nature of intervention

The interventions were heuristics: that is, plans and scaffolds for writing argument or 'devices to teach students a pattern of thought'. These were based on Toulmin's (1958) model of argument, which defined claims or propositions on the one hand, and supporting evidence or grounds on the other, as the basic elements of an argument (this is called 'thesis-support' in the article).

The two particular heuristics were a 'pyramid' closely modelled on Toulmin's model, with the claim or opinion supported by evidence; and by warrants and backing that make the connection between the opinion and the evidence valid. The second heuristic was a 'bridge', based on classical rhetoric (and most immediately on Fulkerson's 1996 work) in which the reason for a position was connected to an opinion by facts, 'if/thens' and values.

Summary of results

Gain scores were higher in the experimental groups than in the control groups as far as argumentative development and voice were concerned, but not significantly higher for conventions. Development gain scores for cultural minorities were significantly higher in the experimental groups, but only marginally higher for the White students. The same is true for the results on voice.

From the survey (questionnaire) results, it appeared that Hispanic- and African-Americans were less aware of the thesis-support model of argumentation than White students; Asian-Americans were excluded from this finding because of the small sample.

The results also suggest that 'the effect of ethnicity is not due to the degree of familiarity with thesis-support argumentation or, alternatively, that the self-report measure of familiarity was inadequate' (p 67).

However, the author suggests that a more balanced sample of white and minority ethnic students is required to confirm these findings.

and other interventions took place in the context of a writing process model (De La Paz and Graham, 2002; Englert et al., 1991; Troia and Graham 2002); some degree of cognitive reasoning training (Englert et al., 1991); peer collaboration modelling a dialogue that was assumed to be internalised as thought (ibid.); self-regulated strategy development (a kind of personal target-setting) (Graham et al., 2005); and the ability to match data against an internal model/schema (Ferretti et al., 2000).

Of the actual interventions, the category listed by the largest number of studies (six) was heuristics or scaffolding devices used by teachers to help students write in argumentative mode. This approach was followed in a number of studies by the use of oral argument to inform argumentative writing (e.g. Englert et al., 1991; Troia and Graham, 2002); by teacher modelling (e.g. De la Paz and Graham, 1997; Troia and Graham, 2002); by the explicit identification of goals and audiences for writing (e.g. De la Paz and Graham, 2005; Graham et al., 2005); and by 'procedural facilitation' or coaching through the process of writing argument (De La Paz and Graham, 2002).

Full details of interventions and summaries of results are contained in the summary tables in Appendix 4.1.

4.5 In-depth review: quality-assurance results

Data extraction was undertaken by four reviewers in the Review Group (RA, GL, NM and CT) and the EPPI-Centre link person (KD). All 16 studies included in the in-depth review were double data-extracted by pairs of reviewers, working independently who then met to discuss any disagreements. Four of the studies were double data-extracted by each member of the Review Group paired with the EPPI-Centre link person. The other 12 studies were data extracted by pairs of reviewers. Finally all sixteen papers were read and the data extraction and WoEs checked by CT and RA, again working independently and then meeting to discuss any inconsistencies.

4.6 Nature of actual involvement of users in the review and its impact

The review was undertaken primarily by researchers, lecturers and an information officer at the University of York, with help from two senior advisory colleagues linked to the DfES National Strategies. These colleagues provided guidance, text and information on the context for the present review, making sure there was relevance in relation to the National Strategies; and providing text and guidance on the implications of the review for policy and practice.

In addition, at protocol and draft report stages, two further groups were involved: PGCE students were involved in assessing the focus and practicality of the design of the review; and policy staff at the DfES were involved in assessing the relevance of the review in relation to the National Curriculum and National Strategies.

4.7 Summary

The review set out to answer the research question 'What is the evidence for successful practice in teaching and learning with regard to argumentative non-fiction writing for 7-14 year olds?'

From a review of the 11 studies summarised above, it appears that certain conditions are either assumed or have to be in place to create a climate for successful practice. These are not specific to argumentative writing but include the following:

- A writing process model in which students are encouraged to plan, draft, edit and revise their writing (De La Paz and Graham, 2002; Englert et al., 1991; Troia and Graham, 2002)

- Self-motivation (in the form of personal target-setting – one aspect of self-regulated strategy development) (Graham et al., 2005)

- Some degree of cognitive reasoning training in addition to the natural cognitive development that takes place with maturation (Englert et al., 1991; Ferretti et al., 2000)

- Peer collaboration, thus modelling a dialogue that (it is hoped) will become internal and constitute 'thought' (Englert et al., 1991)

- Explicit and very clear explanations for students of the processes to be learned

More specifically and more relevantly to the present review, a number of strategies have been identified that have contributed to successful practice in teaching and learning with regard to argumentative writing for 7-14 year olds:

- Heuristics: that is scaffolding of structures and devices that aid the composition of argumentative writing – in particular, planning, which can include examining a question, brainstorming, organising and sequencing ideas and evaluating (De La Paz and Graham, 1997; De La Paz and Graham, 2002; Englert et al., 1991; Troia and Graham, 2002). Planning which is extensive, elaborated and hierarchical can make for more effective argumentative drafting and completion of essays (De La Paz and Graham, 2002). Yeh (1998) used heuristics based on Toulmin (1958) and classical rhetoric.

- The use of oral argument, counterargument and rebuttal to inform written argument (De la Paz and Graham, 1997; Reznitskaya et al., 2001)

- The identification of explicit goals (including audiences) for writing (Ferretti et al., 2000)

- Teacher modelling of argumentative writing (Englert et al., 1991)

- 'Procedural facilitation': that is, coaching by the teaching through the process of writing argument (De La Paz and Graham, 2002).

CHAPTER FIVE
Implications

5.1 Summary of principal findings

In answer to the main research question, 'What is the evidence for successful practice in teaching and learning with regard to argumentative non-fiction writing for 7-14 year olds?' clear results were found. These were that a combination of contextual factors and specific interventions were necessary for successful practice in teaching and learning. The contextual factors were as follows:

- A writing process model in which students are encouraged to plan, draft, edit and revise their writing (De La Paz and Graham, 2002; Englert et al., 1991; Troia and Graham, 2002)

- Self-motivation (personal target-setting as part of self-regulated strategy development) (Graham et al., 2005)

- Some degree of cognitive reasoning training in addition to the natural cognitive development that takes place with maturation (Englert et al., 1991; Feretti et al., 2000)

- Peer collaboration, thus modelling a dialogue that (it is hoped) will become internal and constitute 'thought' (Englert et al., 1991)

The specific interventions were as follows:

- Heuristics: that is, scaffolding of structures and devices that aid the composition of argumentative writing – in particular, planning, which can include examining a question, brainstorming, organising and sequencing ideas and evaluating (De La Paz and Graham, 1997; De La Paz and Graham, 2002; Englert et al., 1991; Troia and Graham, 2002). Planning which is complete, elaborated and hierarchical can make for more effective argumentative drafting and completion of essays (De La Paz and Graham, 2002). Yeh (1998) suggested using heuristics based on Toulmin (1958) and classical rhetoric.

- The use of oral argument, counterargument and rebuttal to inform written argument (De la Paz and Graham, 1997; Reznitskaya et al., 2001)

- The identification of explicit goals (including audiences) for writing (Ferretti et al., 2000)

- Teacher modelling of argumentative writing (Englert et al., 1991)

- 'Procedural facilitation': that is, coaching by the teaching through the process of writing argument (De La Paz and Graham 2002).

Our summary must come with a *caveat*. No single paper suggested all the above, so we are unable to say, at this point, which particular combinations of contextual factors and specific interventions were successful, other than what has been reported in the individual studies: for example, in Englert et al. (1991, one of the most highly-rated studies, the combination is that of a writing process model, some degree of cognitive reasoning training and peer collaboration on the one hand; and the use of 'heuristics' for the improvement of argumentative writing on the other.

5.2 Strengths and limitations of this review

The strengths of the review are that it was undertaken by a team that included researchers, teacher educators, lecturers and policy-makers. It took advice from PGCE students at two points: at the stage of writing the protocol and at first draft stage for the report itself. It was based on the identification of 16 primary studies for the in-depth review. Of these primary studies in the in-depth review, 11 were deemed to be 'medium or above' in terms of weight of evidence/quality, which constitutes a good basis from which to draw conclusions.

The limitations include the fact that it was not possible to answer all the subsidiary questions identified in the review (see section 1.6). There was insufficient basis to be able to say anything conclu-

sive about variation of the evidence with regard to gender, or to pupils with English as an additional language. Studies which were undertaken with pupils with learning difficulties suggested that such pupils would benefit just as much as those without such difficulties, but we do not have a broad enough basis of evidence to be able to generalise from this finding. It must be taken as a suggestion for further research.

Further limitations are evident in relation to the transition from elementary / primary schools to middle schools and thus on to high / secondary schools. None of the included studies in the in-depth review addressed this issue. We *were* able to report a synthesis of studies that suggested aspects of teaching and learning that best helped pupils to improve the quality of their argumentative writing.

Another possible limitation of the review is that, of the 16 studies selected for the in-depth review, 14 were conducted in the US and two in Canada. None was conducted in the UK or in other English-speaking countries around the world. While this is not a limitation in itself, it means that the results must be treated with a degree of caution: what works in a classroom in the US or Canada might not work in the same way in the UK. For example, the assumption in many of the US studies that the 'five-paragraph essay' is behind much of the practice cannot be made about practice in the UK. That having been said, many of the curricular innovations described in the studies would not be unknown to teachers in the UK.

5.3 Implications

5.3.1 Policy

The findings confirm the increased emphasis and secure the place of argumentative writing in the National Curriculum at key stages 2 and 3 in its Curriculum 2000 version – the present version that underpins the curriculum in schools in England and Wales. It was not until the present version that argumentative writing had such a profile within the curriculum; earlier versions downplayed it in relation to narrative, expressive and descriptive writing. The findings also confirm that advances can be made by pupils in the 7-11 age range as well as in the 11-14 age range. There is every reason to believe that the teaching and learning of argumentative writing should start early in Key Stage 2.

In terms of the Primary National Strategy and the Secondary National Strategy, the findings confirm the emphasis that has been put on the process of writing, teacher modelling and peer collaboration in the strategies. The findings also raise interesting questions about critical thinking and cognitive reasoning, where strategies could be developed for improving and challenging pupils' thinking in relation to both argumentative writing and other forms of writing. In this respect, there is a timely connection with the Secondary Strategy's work on think-

ing skills in the 'leading in learning' whole school initiative, and the findings are also significant for the functional skills proposals – both writing and oral work – for developing argument and a concept of progression in teaching and learning of written argument over the key stages. The findings reinforce existing support and guidance on speaking and listening, and may inform future developments in relation to the value of oral argument *per se*, as well as provide a precursor to and preparation for written argument. Furthermore, the findings are useful in identifying the motivational importance of pupils setting, and having choice over, explicit goals for their writing.

Perhaps the key finding in terms of policy is that argumentative teaching strategies cannot be expected to succeed without deep understanding of writing process and its implications for learning; and an encouragement for pupils to work together in solving problems and exploring ideas. Self-motivation and self-regulatory learning strategies are also needed so that the learning is embedded, rather than a superficial response to teaching.

5.3.2 Practice

Further development of practice with regard to the teaching and learning of argumentative writing must take on board what has been said above about the links between contexts for learning and specific 'heuristics' for improving argumentative writing. To use a gardening metaphor, the ground needs to be well prepared for new practices to take root, and for sustained and vigorous growth to take place within a framed ('scaffolded') curriculum plan.

Our knowledge of textbooks and practices in the field suggests that few programmes for teaching argument address both aspects of the problem. The 'critical thinking' movement has spawned a variety of approaches, as have innovations in learning styles and strategies. Neither of these traditions has been linked specifically to the teaching of argumentative writing in English, nor across the curriculum. There has also been little in the way of transfer of argumentative skills across the transitions from primary to secondary schools in the UK.

There is every indication, however, that practitioners and policy-makers working within the context of the National Curriculum for English in England would be receptive to the recommendations made in this report. The genre-based approach to English pedagogy introduced with the National Literacy Strategy in the final years of the 20th century challenged the perceived dominance of narrative within the classroom by encouraging a focus upon so-called non-fictional genre such as 'discursive writing'. The genre-based approach also brought with it an explicit concern not only for the ways in which texts are structured but also for how they seek to position their readers at word, sentence and whole text level.

This change of pedagogical focus required teachers to reposition themselves within the classroom, so that they operated less as facilitators of learning and more as expert practitioners who needed to be skilful exponents of some of the key strategies recommended by this report: for example, modelling good practice as writers themselves or coaching their pupils in the acquisition of explicit writing techniques and strategies. Developments in technology – such as the increasing use of interactive whiteboards with internet access to a wealth of resource materials – have made the explicit modelling and sharing of writing practices a regular and engaging shared experience in many English classrooms.

Most significantly, perhaps, 'argument' is now firmly embedded within the assessment procedures of the English National Curriculum. The writing 'triplet' of *argue, persuade and advise* runs like a binding thread through key stages 3 and 4. At Key Stage 3, for example, several of the Assessment Focuses for EN3 (Writing) examine pupils' ability to attend to 'deep' and 'surface' structural features of their writing, with a particular emphasis upon 'composition and effect' and the ability to gauge requirements of audience and purpose. Typical national examination assignments at key stages 3 and 4 might be to argue a case for the retention of a public park as a recreational space for young people, to write a letter to a headteacher, arguing for a change in a school's curriculum, or to write in role as a character from a play by Shakespeare, urging a particular course of action. For EN1 (Speaking and Listening) assessment, pupils might be encouraged – and again this is an effective strategy highlighted by the report – to work as a team on the creation of a poster designed to argue a particular case.

In terms of 'curriculum backwash', this shift in assessment focus has encouraged a corresponding classroom emphasis upon the structures and strategies associated with argumentative writing. At a lexical and syntactical level, for example, pupils might be taught how to use a 'discursive marker' such as the word 'however' within a sentence. They might be encouraged to learn and consolidate argumentative strategies through the acquisition of mnemonics such as 'a forest': *alliteration, facts, opinion, repetition, emotive language* and *three (rule of)*.

The emphasis upon written argument is not of course confined to the English classroom. The National Strategy has encouraged once again an attempt to involve all teachers in the explicit development of language skills. Argument has an important part to play in the History lesson, for example, or the Science laboratory. Recent initiatives in citizenship education have reinforced the importance of members of a democratic state be able to argue their case or to weigh the arguments of others. Interest in metacognition has been renewed through the development of thinking skills in the classroom and through attempts to help pupils take responsibility for reflecting upon their own learning and achievement.

Practitioners - particularly those new to teaching – need the kind of guidance that this report can give on how to model good argumentative writing practice themselves, on how to coach their pupils in the most effective and proven writing procedures, and on how to establish engaging learning opportunities in which the skills of written argument might be developed and incrementally honed across the key stages and across all four modalities of English.

5.3.3 Research

The systematic review provides an excellent basis for further research.

First and foremost, we recommend the undertaking of new primary studies in the teaching and learning of argumentative writing in the UK. The age-group from 7-14 appears to be an important one for such studies, as this is the period during which argumentation can be developed in writing in preparation for more advanced work from 14-19.

Second, we believe that a number of large-scale trials in primary and secondary and secondary schools might be undertaken to test the worth of different interventions intended to improve the quality of argumentative writing. These could be supplemented or work alongside case studies of classrooms or schools that aim to provide qualitative data on the particular circumstances of teaching and learning such writing.

Third, the suggestion needs to be pursued that pupils with learning difficulties can learn to write better argument, alongside pupils without such difficulties.

Fourth, there need to be closer links between review groups undertaking work on critical thinking and other forms of approach to the improvement of reasoning in school education.

Fifth, international comparative studies would be helpful in determining the national characteristics of the relationships between reasoning and argumentation. It cannot be assumed that practices in one country transfer easily or readily to another.

Finally, this is a field in which traditional research methods can be helpful in examining the effects and the nature of teaching approaches with regard to argumentative writing. However, we should not close the door to new and innovative approaches to research methodology in the field. To help pupils write better argument, we may well need to devise research that gets closer to the heart of the problem.

CHAPTER SIX
References

6.1 References included in map and synthesis

Studies in bold were included in the in-depth review.

Inked papers in the map are indicated with an asterisk(*).

6.1.1 Primary studies

Aulls MW (2003) The influence of a reading and writing curriculum on transfer learning across subjects and grades. *Reading Psychology* **24**: 177-215.

*Burkhalter N (1994). Applying Vygotsky: teaching preformal-operational children a formal-operational task. Unpublished research report. University of Wyoming, WY. ERIC document ED379208.**

*Burkhalter N (1995) A Vygotsky-based curriculum for teaching persuasive writing in the elementary grades. Language Arts 72: 192-199.**

Cox BE, Shanahan T, Tinzmann MB (1991) Children's knowledge of organization, cohesion and voice in written exposition. *Research in the Teaching of English* **25**: 179-218.

Crowhurst M (1990) Reading/writing relationships: an intervention study. *Canadian Journal of Education* 15: 155-172.

De La Paz S (1997) Strategy instruction in planning: teaching students with learning and writing disabilities to compose persuasive and expository essays. *Learning Disability Quarterly* **20**: 227-248.

De La Paz S, Graham S (1997) Effects of dictation and advanced planning instruction on the composing of students with writing and learning problems. *Journal of Educational Psychology* 89: 203-222.

De La Paz S (2001) Teaching writing to students with attention deficit disorders and specific language impairment. *Journal of Educational Research* **95**: 37-47.

De La Paz S, Graham S (2002) Explicitly teaching strategies, skills, and knowledge: Writing instruction in middle school classrooms. *Journal of Educational Psychology* 94: 687-698.

Englert CS, Raphael TM, Anderson LM, Anthony HM, Stevens DD (1991) Making strategies and self-talk visible: writing instruction in regular and special education classrooms. *American Educational Research Journal* 28: 337-372.

Ferretti RP, MacArthur CA, Dowdy NS (2000) The effects of an elaborated goal on the persuasive writing of students with learning disabilities and their normally achieving peers. *Journal of Educational Psychology* 92: 694-702.

Gordon CJ (1990) A study of students' text structure revisions. *English Quarterly* **23**: 7-30.

Graham S, Harris KR, Mason L (2005) Improving the writing performance, knowledge, and self-efficacy of struggling young writers: the effects of self-regulated strategy development. *Contemporary Educational Psychology* 30: 207-241.

*Hallenbeck MJ (1999). *Taking charge: adolescents with learning disabilities assume responsibility for their own writing*. Paper presented at the Annual Convention of the Council for Exceptional Children, Charlotte, NC. ERIC document ED430326.

*Hallenbeck MJ (2002) Taking charge: adolescents with learning disabilities assume responsibility for their own writing. *Learning Disability Quarterly* **25**: 227-246.

35

Hammann LA, Stevens RJ (2003) Instructional approaches to improving students' writing of compare-contrast essays: an experimental study. *Journal of Literacy Research* **35: 731-756.**

Hidi S, Berndorff D, Ainley M (2002) Children's argument writing, interest and self-efficacy: an intervention study. *Learning and Instruction* **12: 429-446.**

Knudson RE (1991) Effects of instructional strategies, grade, and sex on students' persuasive writing. *Journal of Experimental Education* **59: 141-152.**

*Knudson RE (1992). *An analysis of persuasive discourse: learning how to take a stand.* **Paper presented at the Annual Meeting of the National Reading Conference, San Antonio, TA. ERIC document ED353581.**

*Knudson RE (1994) An analysis of persuasive discourse: learning how to take a stand. *Discourse Processes* **18: 211-230.**

Reznitskaya A, Anderson RC, McNurlen B, Nguyen Jahiel K, Archodidou A, Kim S (2001) Influence of oral discussion on written argument. *Discourse Processes* **32: 155-175.**

Sexton M, Harris KR, Graham S (1998) Self-regulated strategy development and the writing process: effects on essay writing and attributions. *Exceptional Children* **64**: 295-311.

Troia GA, Graham S (2002) The effectiveness of a highly explicit, teacher-directed strategy instruction routine: changing the writing performance of students with learning disabilities. *Journal of Learning Disabilities* **35: 290-305.**

VanTassel Baska J, Johnson DT, Hughes CE, Boyce LN (1996) A study of language arts curriculum effectiveness with gifted learners. *Journal for the Education of the Gifted* **19: 461-480.**

VanTassel-Baska J, Zuo L, Avery LD, Little CA (2002) A curriculum study of gifted-student learning in the language arts. *Gifted Child Quarterly* **46: 30-44.**

Yeh SS (1998) Empowering education: teaching argumentative writing to cultural minority middle-school students. *Research in the Teaching of English* **33: 49-83.**

6.1.2 Reviews

Crowhurst M (1990) Teaching and learning the writing of persuasive/argumentative discourse. *Canadian Journal of Education* **15**: 348-359.

Gleason MM (1999) The role of evidence in argumentative writing. *Reading and Writing Quarterly: Overcoming Learning Difficulties* **15**: 81-106.

Newcomer PL, Barenbaum EM (1991) The written composing ability of children with learning disabilities: a review of the literature from 1980 to 1990. *Journal of Learning Disabilities* **24**: 578-593.

6.2 Other references used in the text of the report

Andrews R (1992) An exploration of structural relationships in narrative and argumentative writing, with particular reference to the work of Year 8 students. PhD thesis. Hull: University of Hull.

Andrews R (1995) *Teaching and Learning Argument*. London: Cassell.

Andrews R, Costello P (1992) Improving the Quality of Argument, 7-16 (interim report). Hull: University of Hull, Centre for Studies in Rhetoric.

Andrews R, Costello P (1993) Improving the Quality of Argument, 5-16 (final report). Hull: University of Hull, Centre for Studies in Rhetoric.

Andrews R, Gibson H (1993) A critique of the 'chronological/non-chronological' distinction in the National Curriculum for English. *Educational Review* **45**: 238-249.

Barnes D, Britton J, Rosen H (1969) *Language, the Learner and the School*. Harmondsworth: Pelican.

Bereiter C, Scardamalia M (1982) From conversation to composition: the role of Instruction in the developmental process. In Glaser R (ed) *Advances in Instructional Psychology*. London: Lawrence Erlbaum Associates.

Bereiter C, Scardamalia M (1987) *The Psychology of Written Composition*. Hillsdale NJ: Lawrence Erlbaum Associates.

Britton J (1972) *Language and Learning*. Harmondsworth: Pelican.

Britton J, Burgess T, Martin N, McLeod A, Rosen H (1975) *The Development of Writing Abilities*. London: Macmillan.

Cox B (1991) Cox on Cox: An English Curriculum for the 1990s London: Hodder & Stoughton.

Derewianka B (1990) *Exploring how texts work*. Australia: PETA.

DES (1975) *A Language for Life* (The Bullock Report). London: HMSO.

DES (1989) *English for Ages 5-16*. London: HMSO.

DfEE (1998) *The National Literacy Strategy: Framework for Teaching*. London: DfEE.

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

DfEE (2001a) *Key Stage 3 National Strategy: Framework for Teaching English: Years 7, 8 and 9.* London: DfEE.

DfEE (2001b) *English Department Training 2001.* London: DfEE.

DfEE (2001c) *Literacy Across the Curriculum.* London: DfEE.

DfES (2000) *Grammar for Writing (Key Stage 2).* London: DfES010107/2000, DfES0107/2000V. www.standards.dfes.gov.uk/primary/publications/ literacy/63317/

DfES (2001) *Developing Early Writing.* London: DfES0055/2001, DfES0056/2001. www.standards. dfes.gov.uk/primary/publications/literacy/63337/

DfES (2003) *English Department Training 2003: Improving Writing.* London: DfES.

DfES (2004a) *Literacy and Learning.* London: DfES.

DfES (2004b) *English Department Training 2004: Improving Writing.* London: DfES.

Dixon J, Stratta L (1986) *Writing Narrative – and Beyond.* Sheffield: National Association for the Teaching of English.

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

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

Freedman A, Pringle I (1984) Why students can't write argument. *English in Education* **18**: 73-84.

Goodwyn A (1992) English teachers and the Cox models. *English in Education* **26**: 4-10.

Graves D (1982) *Writing.* Portsmouth, NH: Heinemann.

Hillocks G (1986) Research on written composition: new directions for teaching. Urbana, IL: NCTE.

Hillocks G (1995) *Teaching writing as reflective practice.* NY: Teachers College Press.

Kress G (1994) *Learning to Write.* London and New York: Routledge.

Lewis M, Wray D (2000) *Literacy in the Secondary School.* London: David Fulton.

Moffett J (1968) *Teaching the Universe of Discourse.* Boston: Houghton Mifflin.

Appendix 1.1: Authorship of this review

This work is a report of a systematic review conducted by the English Review Group.

The authors of this report are

Richard Andrews (University of York)
Graham Low (University of York)
Nick McGuinn (University of York)
Alison Robinson (University of York)
Carole Torgerson (University of York)

They conducted the review with the benefit of advice active participation from the members of the review group.

For further information about this review, please contact:

Alison Robinson
Department of Educational Studies
University of York
York YO10 5DD

E-mail: ar31@york.ac.uk
Tel: +44 (0)1904 433462
Fax: +44 (0)1904 433459

For further information about the work of the EPPI-Centre, please contact:

EPPI-Centre
Social Science Research Unit
Institute of Education, University of London
18 Woburn Square
London WC1H 0NR

tel: +44 (0)20 7612 6131
fax: +44 (0)20 7612 6800

email: EPPIAdmin@ioe.ac.uk

Review group

Richard Andrews (University of York)
Graham Low (University of York)
Nick McGuinn (University of York)
Alison Robinson (University of York)
Carole Torgerson (University of York)

Advisory group membership

For this review:

Rob Batho (Secondary National Strategy)
Natalie Parish (Department for Education and Skills)
Wendy Pemberton (Primary National Strategy)
Gerry Swain (Secondary National Strategy)
Victoria White (Department for Education and Skills)

Conflict of interest

There were no conflicts of interest for any members of the Review Group.

Acknowledgements

The EPPI-Centre link person for this review was Kelly Dickson.

Appendix 2.1: Inclusion and exclusion criteria

Inclusion criteria

- Must focus on the teaching and/or learning of argumentative non-fiction writing[1] in English[2]

- Must be teaching and/or learning of English as a first, additional or second[3] (not foreign) language

- Must focus on children or young people aged between 7 and 14 years

- Must be study type B (exploration of relationships), C (evaluation) or E (review) as defined by the EPPI taxonomy of study types (EPPI-Centre, 2002a)

- Must be published or unpublished (but in the public domain) between 1990 and the present

NOTES

1. Non-fiction writing includes writing to inform, explain and describe (reports, explanations, manuals, prospectuses); to persuade, argue and advise (essays, reviews, opinion pieces, advertisements); as well as writing to analyse, review and comment (commentaries, articles, etc.).

2. English implies both English language and English curriculum.

3. English as an additional or second language is used in the sense in which it is commonly employed in UK educational circles, namely to refer to students in the education system of a largely English-speaking host culture, and who, in theory, are immersed in that culture and environment.

Exclusion criteria

Exclusion on scope

One: Not teaching and/or learning of argumentative non-fiction writing in English

Two: Not teaching and/or learning of English as a first, additional or second language

Three: Not children or young people aged between 7 and 14 years

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) Theoretical paper

(h) Position paper

Exclusion on date

Five: Not published or unpublished between 1990 and 1995

Appendix 2.2: Search strategy for electronic databases

1. Educational Resources Information Center (ERIC)

Searched on 10 June 2005 via Cambridge Scientific Abstracts
611 records retrieved

KW=writ* and KW=(non-fiction* or nonfiction*)
or
KW=writ* and KW=(non-literary or nonliterary)
or
KW= writ* and KW=(non-narrative or nonnarrative or discursive)
or
KW=argumentation or (persuasive writing) or (expository writing)
or
KW=writ* and argument*
or
KW=writing and DE=(persuasive discourse) or (outlining discourse)
or
KW=writing and DE=(opinion papers) or essays)
or
KW=writing within5 (article* or review*)
or
KW=writing and KW=(brochure* or prospectus)
and
KW=(child* or pupil* or student*)and(school* or education)
and not KW=(adult* or medical or law or legal or undergraduate or postgraduate) and not DE= (higher edu-
cation) or (science education) or (mathematics education)
or
KW=(key stage) or (national literacy strategy) or (primary school*) or (middle school*) or (secondary school*)
or (elementary school*) or (high school*) or (independent school*) or (public school*) or (primary education)
or (elementary education) or (secondary education)
and not DE=(science education) or (science instruction) or (mathematics education)
and
PT=(072 book /product reviews) or (140 reports general) or (142 reports evaluative) or (143 reports
research) or (080 journal articles) or (010 books) or (040 dissertations /theses) or (041 dissertations /theses
doctoral dissertations) or (042 dissertations /theses masters theses)

Limited to: English only
Date range: 1990-2005

2. PsycINFO

Searched on 10 June 2005 via OVID.
494 records retrieved

KW=writ* and KW=(non-fiction* or nonfiction*)
or
KW=writ* and KW=(non-literary or nonliterary)
or
KW= writ* and KW=(non-narrative or nonnarrative or discursive)
or
KW=argumentation or (persuasive writing) or (expository writing)
or
KW=writ* and argument*
or
KW=writing and (persuasive discourse)
or
KW=writing and KW=((opinion paper*) or essay* or brochure* or prospectus))
or
KW=writing near5 (article* or review*)
and
KW=(child* or pupil* or student*) and (school* or education) and not KW=(adult* or medical or law or legal or undergraduate or postgraduate or (higher education))
or
KW=(key stage) or (national literacy strategy) or MJ,MN=("Elementary-Education") or ("Elementary-School-Students") or ("Elementary-Schools") or ("High-School-Education") or ("High-School-Students") or ("High-Schools") or ("Intermediate-School-Students") or ("Junior-High-School-Students") or ("Junior-High-Schools") or ("Middle-School-Education") or ("Middle-School-Students") or ("Middle-Schools") or ("Primary-School-Students") or ("Private-School-Education") or ("Public-School-Education") or ("Secondary-Education")
and
DT:PSYI = AUTHORED-BOOK or CHAPTER or DISSERTATION-ABSTRACT or EDITED-BOOK or JOURNAL or PEER-REVIEWED-JOURNAL or REVIEW

Limited to: English only
Date range: 1990-2005

3. Social Sciences Citation Index (SSCI)

Searched on 15 June 2005 via Web of Knowledge
318 records retrieved

KW=writ* and (non-fiction* or nonfiction* or non-literary or nonliterary or non-narrative or nonnarrative or discursive or argument* or essay* or brochure* or prospectus)
or
KW=argumentation or (persuasive writing) or (expository writing) or (persuasive discourse)
and
KW=(teach* or school* or education*) not (adult* or law* or medical or science or mathematics or undergraduate* or postgraduate*)

Limited to: English only
Date range: 1990-2005

4. Applied Social Sciences Index and Abstracts (ASSIA)

Searched on 13 June 2005 via Cambridge Scientific Abstracts
29 records retrieved

KW=writ* and KW=(non-fiction* or nonfiction*)
or

KW=writ* and KW=(non-literary or nonliterary)
or
KW= writ* and KW=(non-narrative or nonnarrative or discursive)
or
KW=argumentation or (persuasive writing) or (expository writing)
or
KW=writ* and argument*
or
KW=writing and DE=(persuasive communication)
or
KW=writing and ((opinion paper*) or essay*))
or
KW=writing within5 (article* or review*)
or
KW=writing and (brochure* or prospectus)
and
KW=(child* or pupil* or student*)and(school* or education) and not KW=(adult* or medical or law or legal or undergraduate or postgraduate) and not DE= (higher education) or (science education) or (mathematics education)
or
KW=(key stage) or (national literacy strategy) or DE=("Secondary schools" or "Boarding schools" or "Comprehensive schools" or "Elementary schools" or "Grammar schools" or "High schools" or "Junior high schools" or "Junior schools" or "Junior secondary schools" or "Middle schools" or "Primary schools" or "Private schools" or "Public schools" or "Elementary education" or "Secondary education") and not DE=("science education" or "science instruction" or "mathematics education")

Limited to: English only
Date range: 1990-2005

5. British Education Index (BEI)

Searched on 15 June 2005 via Dialog@site
153 records retrieved

KW=writ? and (non-fiction? or nonfiction? or non-literary or nonliterary or non-narrative or nonnarrative or discursive or argument? or essay? or brochure? or prospectus)
or
KW=argumentation or (persuasive writing) or (expository writing) or (persuasive discourse)
or
KW=writ* within5 (article? or review?)
and
KW=(key stage) or (national literacy strategy) or (child? or pupil? or student? or school? or education) and not KW=(adult? or law or legal or medical or undergraduate? or postgraduate?)

Limited to: English only
Date range: 1990-2005

6.C2SPECTR

Searched on 20 June 2005
22 records retrieved

KW=writ*
and
KW=(non-fiction* or nonfiction* or argument* or persuasive or expository or discursive or essay*)

Limited to: English only
Date range: 1990-2005

APPENDIX 2.3   EPPI-Centre keyword sheet, including review-specific keywords

V0.9.7   *Bibliographic details and/or unique identifier*

A1. Identification of report
Citation
Contact
Handsearch
Unknown
Electronic database (please specify)
....

A2. Status
Published
In press
Unpublished

A3. 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)
....
....
....
....

A4. Language (please specify)
....

A5. In which country/countries was the study carried out? (please specify)
....
....

A6. 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) ....

A7. Curriculum
Art
Business studies
Citizenship
Cross-curricular
Design and technology
Environment
General
Geography
Hidden
History
ICT
Literacy – first language
Literacy further languages
Literature
Mathematics
Music
PSE
Physical education
Religious education
Science
Vocational
Other (please specify) ....
....
....
....
....

A8. Programme name (please specify)
....
....

A9. 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) ....

A10. Age of learners (years)
0–4
5–10
11–16
17–20
21 and over

A11. Sex of learners
Female only
Male only
Mixed sex ....

A12. 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 (please specify)
....

A13. 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

Review-specific keywords

1. Type of text

- Essay (i.e. the generic term for a school assignment involving argumentation)
- Review (including book review)
- Opinion piece
- Advertisement
- Commentary
- Article (e.g. feature article for a magazine that requires claims, evidence, etc.)
- Report
- Other (Please specify.)

2. Length of intervention

- One week or less
- Up to one month
- Between one and six months
- Over six months
- Not stated

3. Nature of intervention

Please give details.

4. Criteria for assessing argumentativeness

- At word level
- At sentence level
- At text level

(Please annotate details as appropriate.)

5. Writing outcomes

- Pre- and post-test results
- Post-test only
- Exam results
- Written work or naturally occurring texts (i.e. not exam or test-based)
- Other (Please specify.)

6. Types of learners

- Upper ability/gifted
- Mixed attainment classes/'normally attaining' pupils
- Pupils experiencing learning disabilities
- Not stated

7. Age/age range

Please state age, age range and grade level

8. English

- As a first language
- As a second/additional language
- Not stated

Appendix 4.1: Details of studies included in the in-depth review

1. Burkhalter N (1995) A Vygotsky-based curriculum for teaching persuasive writing in the elementary grades. *Language Arts* 72: 192-199

2. Burkhalter N (1994) *Applying Vygotsky: teaching preformal-operational children a formal-operational task*. Unpublished research report. University of Wyoming, WY

Country of study
USA

Age of learners
9-11: grades 4 and 6

Type of study
Researcher-manipulated evaluation: controlled trial

Aims of study
To test the notion that preformal-operational fourth graders would improve their ability to perform a formal-operational task, - persuasive writing - with the help of adults and peers (p 192).

Summary of study design, including details of sample
This is a controlled (cluster) trial. We are not informed how the classes were allocated. The experimental group received explicit teaching to six specific objectives related to persuasive writing over a three-week period. There were 80 students in the experimental group and 84 in the comparison classes. More boys (95) were involved in the study than girls (69). It is not clear whether 11 students failed to complete the study.

Nature of intervention
All the fourth- and sixth-grade participants wrote two persuasive essays, with three weeks of instruction intervening for the experimental group. Instruction consisted of daily 45-minute lessons, involving brainstorming, supporting, conferencing, and editing on different topics. Objectives of the intervention were to help motivate students to organise a persuasive essay; identify and anticipate a reader's objections; transfer oral argumentation to written; and to gain instruction in written argument, after debate. All the participants wrote five 'compositions' during the period of the study. Two of

these formed the pre- and post-test essays. Detailed accounts of the interventions are contained in the article (pp194-195).

Data-collection instruments, including details of checks on reliability and validity
The pre- and post-test assignments asked the participants to write a persuasive letter to the school principal. These pieces of persuasive writing were used to measure the participants' ability to deploy 'claims', 'data' and 'warrants' (Connor, 1990; McCann, 1989) in their writing. The treatment groups were deliberately not randomised in case 'the children might have behaved differently had they known they were in a special study' (p 194). The pupils in the experimental group were not told that they were being taught a curriculum that was different from that of the control group. 'All experimental classes followed the normal process approach for writing tasks' (Paper Two, p 12).

Methods used to analyse data, including details of checks on reliability and validity
The scale adapted from Connor (1990) and McCann (1989) was modified for use with young writers (p. 196) and informed by the work of Toulmin (1958 and 1984) - (see also Paper Two). Separate ANOVAs were performed to determine whether there were differences in the post-test persuasive essay scores for three dependent variables: claims, data, and warrants. Possible scores for each trait were 0, 2, 4, or 6. The independent variables were gender (male, female); treatment (comparison, experimental); and grade level (4, 6) (p 196). The compositions were rated anonymously by three graders. All compositions were graded at the completion of the instruction. Cronbach alpha coefficients suggested that inter-rater reliability was quite high (Paper Two, p 11).

Summary of results
The author reports 'three major findings' (p 197):
   (1) All experimental students scored significantly higher on the post-test than those in the comparison group, who had no instruction in persuasive writing. In addition, even though fourth-grade experimental students scored below the sixth-grade experimentals, their improvement was significant, thus suggesting that children as young as nine can improve their ability to write persuasive essays, even though it is very hard for them.
   (2) All girls had higher pre- and post-test scores than boys.

(3) The most difficult aspect of persuasive writing for all students was warrants.

Conclusions
The authors conclude that their findings support a dynamic, Vygotskian approach to teaching and learning – one which triggers children's potential through adult assistance rather than a more rigid one which bases curriculum design on what children are or are not capable of doing (p 198).

Weight of evidence A (trustworthiness in relation to study questions)
Low to medium: It is not clear how the classes were allocated. The sample is rather small for a trial with a cluster design and no attempt has been made to calculate an 'effective sampling size'. Not all the data has been presented. The control group is inappropriate for investigating the effectiveness of the intervention as it does not receive any alternative intervention. No account has been taken of the clustered nature of the data in the analyses. The study appears to be biased towards Vygotsky and away from Piaget.

Weight of evidence B (appropriateness of research design and analysis)
Low: A controlled cluster trial is potentially an appropriate method. However, this one is weakened by a lack of contextual and procedural information regarding the nature and allocation of the sample and by a failure to take into account the clustered nature of the data in the analyses.

Weight of evidence C (relevance of focus of study to review)
Medium: The study provides useful information about the treatment group's ability to deploy 'claims', 'data' and 'warrants' in persuasive writing. However, the sample, although reasonably large for an individual trial, is not so large for a trial with a cluster design. Not all the data has been presented. The control group is inappropriate for investigating the effectiveness of the intervention.

Weight of evidence D (overall weight of evidence)
Low: The study as a whole can only be graded 'low' on account of the limitations stated. Essentially, these are its sampling techniques, the lack of contextual detail, and the lack of detail in the description of the actual interventions.

Crowhurst M (1990) Reading/writing relationships: an intervention study. *Canadian Journal of Education* 15: 155-172

Country of study
Canada

Age of learners
11-12: Canadian grade 6

Type of study
Researcher-manipulated evaluation: randomised controlled trial

Aims of study
To discover whether students' 'writing of persuasion' (p 157) can be improved by instruction, and specifically whether practice in reading improves writing and vice versa.

Summary of study design, including details of sample
Students in two sixth-grade classes in each of two schools (N=104) were stratified by sex and ability and randomly assigned to one of the following groups for
1. instruction in a model of persuasion with writing practice
2. instruction in a model of persuasion with reading practice
3. practice in reading, but no instruction in the model
4. group discussion activities (control group)

Pre-tests and post-tests required subjects to write a recall protocol for a persuasive reading and to write two persuasive compositions. Instruction comprised: 10 sessions (2 sessions x 5 weeks) of 45 minutes, 3 instruction groups, 1 control.

There were two classes in each of two schools (N=110).

Nature of intervention
Students were divided into four groups to examine different combinations of input/skills:

Writing instruction
Instruction = provision of a model structure, collaborative brainstorming, draft revision in pairs, + teacher feedback of four 'pro' and 'con' essays.

Reading + instruction
Students saw the writing structure model. Instruction emphasised the readings' structure and content. They read five specially-written paired pro-con texts and discussed them - including identifying the structural elements.

Reading + discussion
Instruction was in group discussion skills. They read the same texts as Group 2.

Control (discussion)
Instruction in group discussion. Students did group tasks.

Data-collection instruments, including details of checks on reliability and validity
The baseline reading scores were presumably taken from school records (no procedural details reported).

Reading pre-test: 45 minutes, reading a text (one of a pool of two), then recalling all information in it.

Writing pre-test: 45 minutes, write two essays, each on a separate day. The order of topics for reading and writing tests was counterbalanced (p 159).

Writing post-test: 45 minutes 45 minutes, write two essays, each on a separate day. The order of topics was counterbalanced (p 159).

Essays were rated by four experienced sixth-grade teachers. Topics were sequenced for rating so as to counterbalance them. Inter-rater reliability for essay quality (using Ebel 1979) ranged from 0.95 to 0.92.

The reading passages were specially written at an appropriate reading level (Dale-Chall raw scores: 5.83 and 6.08).

Methods used to analyse data, including details of checks on reliability and validity
Means and SDs for dependent measures (Table 2)

ANOVA for writing quality scores - giving differences between post-test compositions of the writing and reading and instruction groups and the control group

Raters were entered individually (as part of a 'Judge' variable) into the ANOVA with Test and Group.

There is an implied choice of which follow-up planned comparison test to use.

Summary of results
Reading: No significant differences found.
Writing quality: 1. Significant main effect for test (p<0.001) and text x group interaction (p<0.01). At

pre-test, reading + instruction group scored below the control group. At post-test, the writing and the reading + instruction groups scored higher than the control group. This confirmed hypotheses 1a and 1b but not 1c. 2. Significant main effect for judge; 'considered not to be significant since the effect was evenly distributed across groups and tests' (p 161). 3. ANOVAs on specific features of the essays of the writing, reading+instruction and control groups showed: number of reasons/100 words: significant main effect for Test (p<0.001). The number of reasons was generally smaller for the post-test than the pre-test; number of elaborations/100 words: significant main effect for Test (p<0.05) and Test x Group interaction (p<0.001). 4. Writing group used fewer elaborations than the control group at pre-test, but more at post-test; number of conclusions: significant main effect for Test (p<0.001) and Group (p<0.01), plus significant Test x Group interaction (p<0.001). No differences at pre-test, but at post-test, the writing group used more conclusions.

Conclusions
1. 'the persuasive writing of students in the upper elementary school can be improved by instruction. Two of the three instructional groups … scored significantly higher than the control group on writing quality on the post-test' (p 164).

2. 'instruction in the (writing) model plus either writing practice or reading practice tended to result in fewer, more elaborated reasons and in more conclusions' (p 164).

Crowhurst argues that, while differing text length between writing and control texts might have partly influenced raters' judgements of holistic quality, the subsequent finding of significant feature differences rules this out as a major explanation of treatment effect.

3. Instruction in the writing model may have aided students to 'organize what they had to say'. Two examples are cited.

4. Cross modal effects: No evidence of writing affecting reading. Crowhurst argues that this might be because the reading tests were too long and difficult (no validation work beyond establishing Dals-Chall reading age was reported)

The evidence of reading affecting writing is unclear.

The reading + instruction group made great gains from pre-test to post-test, but there is no way to establish how far this was due to instruction in the writing model, or to establishing a persuasion schema while reading/discussing.

It was not clear how far the reading + discussion students actually read the texts, so it was not possible to measure the impact of reading.

Weight of evidence A (trustworthiness in relation to study questions)
Medium: The lack of baseline data, test validation, instruction validation/monitoring and writing feature reliability checks make the trustworthiness medium.

Weight of evidence B (appropriateness of research design and analysis)
Medium: Although this was an RCT, the lack of baseline data, test validation, instruction validation/monitoring and writing feature reliability checks which make the trustworthiness medium for the study objectives make it medium for the review question.

Weight of evidence C (relevance of focus of study to review)
Medium: The general focus is highly relevant to the review: teaching persuasive writing in the light of schema theory. The sampling and the measures (v. WOE B comment) reduce it to medium.

Weight of evidence D (overall weight of evidence)
Medium

De La Paz S, Graham S (1997) Effects of dictation and advanced planning instruction on the composing of students with writing and learning problems. *Journal of Education Psychology* 89: 203-222

Country of study
USA

Age of learners
9-14: grades 5-7

Type of study
Researcher-manipulated evaluation: randomised controlled trial

Aims of study
To examine the effects of explicit instruction in advanced planning on the dictation and writing of opinion essays by students with LD.

Summary of study design, including details of sample
This is an individual RCT with allocation into one of four groups. Pre-, post- and maintenance tests. (N=42)

Nature of intervention
Students were taught a validated strategy for developing, evaluating and organising ideas for their essays in advance of composing. Students were encouraged to modify and supplement this initial plan (as needed) while composing. Students assigned to the comparison condition received no instruction in advanced planning, but rather they learned about the characteristics of good essays for meaning and structure, and composed and shared their own essays with peers.

A unique feature of the study involved the use of dictation as a means for generating not only the essay but also the students' advanced plans.

Data-collection instruments, including details of checks on reliability and validity
De La Paz scored all essays; a random sample (25%) was scored by another rater.

'Planning and composing times were measured by individual instructors during testing; reliability data were not available. Length of essays was determined by computer; reliability was not necessary. Inter-rater reliability was calculated using a Pearson product-moment correlation for all dependent measures that were scored on a continuous scale. Percentage of agreement was calculated for dependent measures that were scored on a categorical scale (i.e. when ratings resulted in a score of 1 or 0) because the Pearson correlation is not accurate for these scores. Criterion for inter-rater reliability scores when training was at least 80%' (p 210).

A randomly selected 1/3 of plans were additionally rated 'by a rater unfamiliar with the purpose and design of the study. Inter-observer agreement between the two rates (Pearson p-m correlation) was 0.99' (p 210).

Transformations: As with plans. Correlations were: total: 0.93 (range: 0.075-0.9 for the separate categories).

Essay elements: As with plans and transformations, but using 1/4 of the essays. Correlations: Total: 0.98 (range: 0.86-0.94 for specific elements).

Essay coherence: All essays re-rated by independent rater, r=0.95.

Essay quality: 'Two regular education sixth-grade teachers who were unfamiliar with the design and purpose of the study independently scored all essays' (p 210). r=0.8, with differences resolved by negotiation.

No details of validity.

Methods used to analyse data, including details of checks on reliability and validity

Means and SDs of plans and essay scores

Also ANOVA and ANCOVA using pre-test as covariate

Fisher-Hayter procedure with covariate adjustment

Johnson-Netman procedure (when assumptions of F-H not met)

Hotelling's T squared

One-way ANOVA

Percentages, means and SDs

Reliability and validity checks not applicable - standard statistical techniques

Summary of results

The combination of dictation and instruction in advanced planning resulted in more complete and qualitatively better essays in contrast to those written by students in the comparison condition on both a post-test and a two-week maintenance probe.

At post-test and maintenance, students who received advanced planning instruction spent more time planning than students in the two comparison conditions.

Students assigned to the two advanced planning conditions did not differ significantly in the number of propositions in their plans at post-test or at maintenance.

Transformations: Data incomplete?

Advanced planning:

'At post-test and maintenance, students in the two comparison (= control) conditions typically spent less than 30 seconds planning in advance of composing' (p 212).

'The students in the AP condition who wrote spent more time planning than those in the CD condition ... [or] CW condition'.

'...students in the AP condition who dictated spent more time planning than students in the CW ... and CD conditions.'

'Therefore, at post-test and maintenance, students who received advanced planning instruction spent more time planning than students in the two comparison conditions'

(All p 213)

Propositions: '...students assigned to the two AP conditions did not differ significantly in the number of propositions in their plans at post-test or maintenance (one-way ANOVA post-test p=0.149; maintenance: p=0.491)

Transformations:The two advanced planning conditions were not significantly different as post-test (p=0.876) or maintenance (p=0.718).

Essay length:

'...students in the AP condition who dictated composed longer essays than students in the CW condition.'

'...students with essays longer than 40 words in the comparison between the two dictation conditions benefited significantly more from the AP condition than the comparison instruction.'

'...students with essays longer than 43 words at pre-test benefited significantly more from the use of dictation at post-test than they did from the use of writing.'

'No significant differences were found between students in any of the four instructional conditions on the maintenance probe.'

(All p 214)

Essay elements: 'essays composed by students in the planning condition who dictated contained more func-

tional elements than students in the CW condition at post-test ... and at maintenance' (pp. 214-5)

Essay coherence: Post-test 'students in the AP and dictation condition composed more coherent essays than students in the CW condition'

Maintenance:

'...students in the CD condition composed more coherent essays than students in the CW condition.'

'...students with coherence pre-test ratings below a score of 2.0 benefited significantly more from the combination of advanced planning and dictation than the combination of comparison instruction and writing. Furthermore, in the comparison of the two writing conditions, students with coherence pre-test ratings below a score of 2.1 benefited significantly more from advanced planning instruction than they did from the comparison instruction. Students with high coherence ratings were not significantly different however.'

(All pp 214-215)

Essay quality: Post-test

AP students who dictated composed qualitatively better essays than CW students.

Maintenance: AP students (writing or dictation) composed qualitatively better essays than CW students.

'at maintenance, for students with pre-test quality ratings below 2.7 in the contrast between students in the AP condition who wrote to students in the CD condition , students benefited significantly more from comparison instruction and dictation than they did from instruction in AP and writing. In addition, students with pre-test quality ratings above 3.2 in these two conditions benefited significantly more from advanced AP instruction and writing than they did from comparison instruction and dictation.

When comparing students in the AP condition who dictated to students in the CD condition, students with pre-test quality ratings above 3.3 benefited significantly more from the instruction in AP and dictation than they did from the comparison instruction and dictation. Students with low pre-test quality ratings in these two conditions were not significantly different, however.

Finally, for students with pre-test quality ratings below 3.0 in the contrast between the two comparison conditions, students benefited significantly more from the use of dictation than they did from the use of writing. Students with pre-test quality ratings above 3.6 in these two conditions benefited significantly more from the use of writing than they did from the use of dictation' (p 216).

Rate: Post-test

CD students composed faster than AP writing students, or CW students (i.e than students who wrote).

Maintenance: CD students composed faster than AP writing students.

AP students composed faster via dictation than writing.

Strategy use: 95% of AP students used the planning strategy at post-test and 91% did at maintenance.

Student evaluations: 90% of dictation students were enthusiastic about dictation.

'Only 4 students ... stated disadvantages to dictation.'

The comments suggest possible limitations on using adults as scribes.

2/3 of AP students liked learning AP strategies.

'Most students indicated that the strategy was easy to learn' (p 218)

Approximately 50% of comparison students 'commented that they had learned how to write better essays' (p 218).

Conclusions

1. As predicted, 'the combination of dictation and advanced planning instruction had a positive effect on the composing of students who find writing and learning challenging' (p 220). However, the prediction that AP students would compose faster was not fulfilled.

2. Dictation had 'weak and inconsistent effects' particularly with comparison students (p 219).

No significant dictation-writing speed or planning differences found at maintenance. Dictation-based essays, were however, more cohesive at maintenance.

Dictation helped students with the lowest quality essays improve the quality (at maintenance), but students with higher quality essays did better when they used writing.

3. 'the impact of a particular writing intervention depends on the child's initial capabilities' (p 220)

This is seen in the aptitude x treatment interventions (e.g. above).

Students with longer pre-test essays benefited more from planning instruction when they dictated.

Students with higher quality essays at pre-test wrote better essays if:

AP involved dictation or writing, but comparison instruction involved writing.

4. 'the writing strategy … helped poor writers do several things they normally do not do' (p 220) (i.e. becoming 'more planful and reflective' from a state of minimal or non-existent planning).

5. Comparisons with 'normally-achieving peers' are needed.

Weight of evidence A (trustworthiness in relation to study questions)

High: The study is an RCT and attention to the validity and usefulness of both instructional conditions was good, plus attention to reliability and validity of the analytic procedures was generally good.

However, the sample size was small.

Weight of evidence B (appropriateness of research design and analysis)

High to medium: The essays are persuasive opinion essays and the study is an RCT. The analytic procedures are also appropriate.

Weight of evidence C (relevance of focus of study to review)

High to medium: The sample is just LD students, so there is no comparison with other children, but otherwise it is highly relevant to the study.

Weight of evidence D (overall weight of evidence)

High to medium: There are no major flaws in the design, execution, analysis, or interpretation.

De La Paz S, Graham S (2002) Explicitly teaching strategies, skills and knowledge: writing instruction in middle school. *Journal of Educational Psychology* 94: 687-698

Country of study
USA

Age of learners
12-14: grades 7 and 8

Type of study
Researcher-manipulated evaluation: randomised controlled trial

Aims of study

'The present study examined the effectiveness of an instructional program designed to improve the writing performance of (American) middle school students' (p 687).

Summary of study design, including details of sample

This is a randomised controlled trial: N = 58 (30 in experimental group and 28 in control group).

Experimental group received SRSD: training in planning, drafting and revision, wrote two collaborative essays, three individual.

Control group wrote one class essay, four individual.

All students selected one essay for class portfolio.

Post-test: plan and essay

Four-week delayed post-test (maintenance probe)

Interviews with teachers about instructional effectiveness

Nature of intervention

The programme primarily focused on teaching students strategies for planning, drafting and revising text. The knowledge and skills needed to support these processes were also emphasised.

The key element of the instructional programme was a strategy that organised and directed the processes for planning and writing an essay. The students developed a plan in advance of writing that involved analysing the demands of the writing assignment, setting goals for writing, and generating and organising material to write about.

With SRSD, students are explicitly taught writing strategies along with procedures for regulating these strategies and the writing process. The procedures include goal setting, self-monitoring, and self-instruction.

Data-collection instruments, including details of checks on reliability and validity

Essays (pre-test, post-test, maintenance): 35 minutes for planning and writing an essay. Expository essays were elicited that involved explanation and persuasion. Three topics were selected from a pool of five. To control for confounding due to differences in student interest, familiarity or knowledge, the administration of the three essay topics were counterbalanced across students and probes using a Latin square design.

Interviews: No method details reported

Details of reliability

For planning: The first author scored all the plans on a five-point scale. A former middle school teacher independently scored a random sample of 20% of the plans - inter-rater reliability was 0.81.

For length of essays: A graduate student counted the number of words in all the essays. A second graduate student independently scored a random sample of 20% of the essays. The inter-rater reliability score was 0.96.

For vocabulary: The first author scored all the essays for maturity of vocabulary. A former teacher scored a random sample of 20% of the essays. Inter-rater reliability was 0.87.

For quality: Traditional holistic rating scale. Two former middle school teachers independently scored all essays. Anchor points for scoring were established by indicative essays produced by students not in the study. The Pearson product-moment correlation for quality was 0.87.

Details of validity:

The pre-test was preceded by an explanation of the sort of text, strategies and style was being sought.

Three topics/prompts were selected from a pool by 'two graduate students, both previous teachers' in terms of interest and difficulty (p 690).

The topics/prompts were counterbalanced across students and probes (using a Latin square design (p 691).

Methods used to analyse data, including details of checks on reliability and validity

A series of 2 (group) x 2 (trial) repeated measures ANOVAs.

In all cases the trials were the post-test and maintenance (not the pre-test).

No details of interview data analysis reported.

No details of reliability or validity.

Summary of results

In comparison with peers in the control condition, students in the experimental condition produced essays that were longer, contained more mature vocabulary, and were qualitatively better. At post-test, the plans of students in the experimental condition were better developed.

Planning: There was a significant main effect for trials (p<0.01) and instructional condition (p=0.00; effect size effect size: 1.17; maintenance effect size: 1.04).

80% of students made no plan on the pre-test; by post-test 90% of experimental students and 30% of control students made accurate detailed plans.

Word length: There was a significant main effect for instructional condition (p<0.01; post-test effect size: 0.82; maintenance effect size: 1.07). Experimental students wrote longer essays.

Vocabulary: There was a significant main effect for trial (p=0.04) and instructional condition (p<0.01; post-test effect size: 1.13; maintenance effect size: 0.94). Experimental students used more long words.

Quality: There was a significant main effect for instructional condition (p=0.00; post-test effect size: 1.71; maintenance effect size: 0.74). The experimental students wrote tests of higher overall quality.

Conclusions

'As expected the writing program had a positive effect on the writing performance of the participating middle school students. Immediately following instruction, students in the experimental group produced essays that were longer, contained more mature vocabulary, and were qualitatively better than the essays generated by youngsters in the control classrooms. These effects were maintained on an essay written 1 month after instruction ended.' (p 695)

'... the impact of the experimental treatment was strong, as effect sizes for the three writing product measures ... ranged from 0.82 to 1.71 on the post-treatment and maintenance writing probes' (p 696).

'As with the writing-product measures, the impact of the experimental treatment on students' written plans was quite strong, as effect sizes were greater than 1.00 on both the post-treatment and maintenance writing probes' (p 696).

The authors hypothesise that gains were due primarily to the explicit strategy instruction and to the 'writing skills and knowledge students were taught' (p 696).

The SRSD procedure, with its direct teaching of writing strategies, thus works effectively in traditional classrooms, with regular teachers and 'normally developing' children. Teachers were able to deliver SRSD with 'a high degree of fidelity' (p 696).

Weight of evidence A (trustworthiness in relation to study questions)

Medium: Exclusion of an unknown number of pupils and one cluster after randomisation, and (final) sample was fairly small. The aim was to test the feasibility of SRSD in traditional classrooms, with 'normally-developing' children and regular classroom teachers. The study indicated with clarity that it could be feasible.

Weight of evidence B (appropriateness of research design and analysis)

Medium: The study is good in that it involves direct teaching of explanation and persuasion strategies, skills and knowledge. The study is an RCT, though the rationale for selecting the schools is not provided explicitly, but the unknown exclusions and lowish sample size bring it down to medium.

Weight of evidence C (relevance of focus of study to review)

High: The sample is the right age and the focus is direct teaching of expository writing skills (in the context of a five-paragraph essay). The study focuses on the feasibility of teaching the SRSD procedure by regular classroom teachers. The assessment measures are sufficiently controlled and the assessment measures cover planning, development, length, 'mature' vocabulary and overall quality.

Weight of evidence D (overall weight of evidence)

Medium to high: Beyond the sampling and sample size problems, there are no serious flaws in the study.

Englert CS, Raphael TM, Anderson LM, Anthony HM, Stevens DD (1991) Making strategies and self-talk visible: writing instruction in regular and special education classrooms. *American Educational Research Journal* 28: 337-372

Country of study

USA

Age of learners

9-11: grades 4 and 5

Type of study

Researcher-manipulated evaluation: randomised controlled trial

Aims of study

To evaluate the effectiveness of an instructional intervention that developed students' knowledge of the writing process and the role of expository text structures by making the strategies and contextual constraints explicit ('visible') via dialog with the teacher, solving problems via scaffolding from teacher or collaboratively from/with peers (p 339, p 342)

Summary of study design, including details of sample

This was an individually randomised RCT in which additional comparison groups were formed within the exp and control groups (LA and HA in the regular classroom groups). Pre- and post-tests of writing ability, comprehension and meta-cognitive awareness were administered. Instruction lasted 7 months (October - April).

128 regular + 55 learning disabled = 183

Nature of intervention

The intervention involved training in planning, organising, writing, editing and revising different text types. Work was individual and collaborative, and guided by a series of think-sheets. Assessment involved testing (1) metacognitive knowledge about the writing process, (2) skill at writing explanation and comparison essays, and (3) ability to transfer skills to reading and comprehension of expository text. (p 348).

The treatment group received systematic instruction in text analysis, modelling the writing process, guided practice and independent use of strategies (p 350). They wrote a class paper, an individual paper, then a paper for a class book (p 352).

The control group received regular writing instruction, which involved some planning, brainstorming and collaboration, but was not systematic and (we infer) the ideas were not made 'visible'. They wrote texts 2-3 times a week (p 352).

Assessment in writing and reading took place in the month before and after the instruction. The order of tests was counterbalanced, except that reading tests preceded writing tests (p 349).

Data-collection instruments, including details of checks on reliability and validity

Metacognitive knowledge was elicited by a questionnaire. Writing was tested by requiring passages to be written in class. Comprehension was tested in class.

LD children were tested in small groups; regular children in whole class units. Special education students had the passage read aloud to them (additionally) (p 350). All students were directed to reread the passage as many times as they wished (p 350). The order of the three text structure papers (comparison/contrast, explanation and expert) was counterbalanced across classrooms at the same grade level. Adequate time was given for students to complete the metacognitive questionnaires. Reliability of the metacognitive questionnaire scoring was reported as 98%. For writing, the scorers were blind to the assignment of subjects. Each student's written composition was read independently by two coders who assigned four scores per paper. Reliability was scored on 10% of the composition measures. Reliability was calculated by dividing the number of agreements by the sum of agreements plus disagreements. On all variables, reliability was above 80%. Reading tests were administered before writing tests. They were written at two different reading levels to ensure students could read the passage without difficulty. LD students had the passage read aloud before reading it in text form, to avoid problems with the mechanics of reading and to minimise differences between regular and LD groups due to reading fluency or lack of experience with expository prose. The order of the writing tests was counterbalanced across classrooms. LD students could ask for a helper to write for them. All students were told not to focus on spelling and punctuation. If a script was not legible, the writer was asked to read out what s/he had written. The questions in the metacognitive questionnaire were delivered orally and read twice, and repeated where necessary. Adequate time was given for all students to complete.

Methods used to analyse data, including details of checks on reliability and validity

1. Essays and reading recall texts graded and questionnaire scores were summed.

2. Multivariate overview was undertaken, controlling for pretest scores (MANCOVA or ANCOVA).

3. Univariate ratios examined where factors yielded significant findings.

4. Where univariate ratios were significant, Scheffe post-hoc comparisons were employed (alpha= 0.05).

Writing texts were rated on four dimensions:

1. Primary trait (use of suggested genre structure and lexis)

2. Holistically (Was the paper interesting? Did it communicate its ideas?)

3. Productivity (number of ideas; ideas were defined differently for each of the three texts (p 354)).

4. Reader sensitivity (text was interesting, explained purpose, showed evidence of reader-writer interactivity)

Details are reported with respect to scoring and coding, but not for the analyses of covariance.

Adjusted means were used at post-test.

Validity was addressed by checking that the alpha level was adjusted appropriately where multiple T-tests were carried out.

Summary of results

Significant main effects for treatment, group and text, and a significant interaction between group and treatment but no other significant interactions. The treatment effects seemed to be attributable to gains in students' holistic ratings, primary traits, and sensitivity to the readers. All these differences favoured the CSIW treatment, suggesting that students in the CSIW treatment showed increasing mastery of the structures underlying text, as well as growing sensitivity to their audience and purpose for writing.

Conclusions

The findings suggested that the dialogic instruction was effective in (a) promoting students' expository writing abilities on two text structures and (b) leading to improved abilities on a near transfer activity, in which students wrote a text structure not taught during the intervention.

Weight of evidence A (trustworthiness in relation to study questions)

High: The results are generally interpretable in terms of the research questions, especially with respect to the writing tasks, and the concrete conclusions drawn by the researchers seem valid. Where the conclusions are speculative, these are flagged as such.

Weight of evidence B (appropriateness of research design and analysis)

High: The study was an RCT with adequate sampling and most important variables counterbalanced.

Weight of evidence C (relevance of focus of study to review)

High: The comparison and explanation tasks are both key examples of non-narrative writing. The broad sampling across the ability range means that it is possible to generalise reasonably confidently. The intervention employs features like collaborative discussion which have been shown to be important in language-related learning.

Weight of evidence D (overall weight of evidence)

High

Ferretti RP, MacArthur CA, Dowdy NS (2000) The effects of an elaborated goal on the persuasive writing of students with learning disabilities and their normally achieving peers. *Journal of Educational Psychology* 92: 687-698

Country of study
USA

Age of learners
9-12: grades 4 and 6

Type of study
Researcher-manipulated evaluation: randomised controlled trial

Aims of study
'... to investigate the effects of giving students an elaborated goal that included explicit subgoals based on the

elements of argumentation as compared with a general goal to convince an audience to agree with their opinion' (p 695).

Summary of study design, including details of sample

Randomised controlled trial
   1. Gathering baseline data and allocation to conditions
   2. Intervention (no details of any instruction are reported)
   3. Students were asked on two occasions to write letters
   - Writing time: 45 minutes
   - No report of the length of time between the occasions.
   - Order of topics balanced within grades and disability status
   N = 124
   General goal = 62
   Elaborated goal = 62
   In each condition Grade 4 = 30, Grade 6 = 32
   In each grade/condition group 50% = LD

Nature of Intervention

'Students in the general goal condition were asked to take a position and write a letter to persuade an audience to agree with them.

   Students in the elaborated goal condition were given the same general goal plus explicit subgoals based on the elements of argumentative discourse.
Subgoals directed students to include (a) a statement of their belief, (b) two or three reasons for their belief, (c) examples or supporting information for each reason, (d) two or three reasons why others might disagree, and (e) why those reasons were wrong' (p 696).

Data-collection instruments, including details of checks on reliability and validity

'On each occasion the prompt was distributed to students and read aloud by the research assistant. Students were asked to read the prompt as well. The students were given 45 minutes to write their letters. Writing samples were gathered from all normally achieving students present on the days data were collected. The authors randomly selected normally achieving students writing samples equal in number to the students with learning disabilities within each grade and goal combination and scored those samples.' (p 696)

   The order of topics was balanced 'within grades and disability status' (p=0.696).

   Essays were scored for persuasiveness by two graduate students who were not otherwise involved in the study. The inter-rater reliability was r=0.72 (homework topic) and 0.73 (violence).

   Essays were scored for argument elements by two pairs of graduate students who were not otherwise involved in the study. A scoring manual was provided and the scorers were trained to use it. Each pair scored a randomly selected half of the papers. A randomly selected 20% of each half was rescored by the other pair.

   Inter-rater reliability for total functional elements r=0.85.

   The coefficient for one functional element (as a percentage of exact agreement) was below 0.80: specifically, elaborations at 0.64.

   Validity was addressed by reading the topics aloud to students, as well as asking them to read the prompts.

Methods used to analyse data, including details of checks on reliability and validity

Authors developed their own primary trait scoring guide to measure overall persuasiveness. Papers were rated on a scale ranging from 0-7 (p 696).

Papers were also scored for elements of argumentative discourse with the use of a scoring manual which provided definitions and examples of each element of persuasive writing (pp 696-697).

   Analysis of variance procedures were used to analyse the data generally.

   No details are reported of reliability of data analysis (beyond ANOVAs examining the relation between IQ and LD).

   Use of Fisher's test of exact probability was used to address validity.

   There was an appropriate use of one- and two-tailed tests.

   Overall, descriptive and statistical analyses are applied appropriately to the data.

Summary of results

Persuasiveness: There was a significant grade by goal effect 'because the elaborated goal enhanced the writing of the sixth-grade students but not that of the fourth-grade students….In addition, normally-achieving students wrote more persuasively than did students with LD, and papers about homework were more persuasive than those about violence' (p 697)

   Elements of argument: There was a significant interaction between grade and goal, attributed to the same reason as for persuasiveness. There was also a significant interaction between grade and disability status, because students with LD included fewer elements than normally achieving students at Grade 4 (but not Grade 6).

   There were significant main effects for goal and disability status. It was found that more elaborated goal students than general goal students gave more than two reasons, plus one or more alternative propositions, alternative reasons and rebuttals. More normally achieving students gave (a) two or more reasons with the homework topic than did the LD students and more normally achieving students gave three or more elaborations than did the LD students and an elaboration with the violence topic.

   Predicting persuasiveness from argument elements: 39% of the variance of the homework topic was predictable from propositions, elaborations and alternative reasons. 44% of the variance of the violence topic was predictable using propositions, reasons, elaborations, conclusions alternative reasons and rebuttals.

Conclusions

1. Results support the position that 'providing older students (i.e. Grade 6) with an elaborated goal led them to include more argumentative elements in their essays and, as a result, to write more persuasively' (p 699).
   2. 'a substantial amount (40%-45%) of the variance in persuasiveness of all students' essays could be accounted for by the inclusion of argumentative elements' (p 699).
   3. (1) and (2) are consistent with the idea of writing as a goal directed activity. 'It appears that normally achieving students and those with LD may benefit from instruction in goal setting' (p 700).
   4. Grade 4 students wrote equally persuasively in both conditions and included equal numbers of argument elements in both essays.
   5. This difference does not seem due to developmental differences, as a comparable proportion of Grade 6 and Grade 4 students attempted to state alternative positions and alternative reasons.
   6. It may be that the combined effects of writing and meeting the Elaborated Goal subgoals exceeded the younger children's executive capacities.
   7. It might also have been the case that the Grade 6 students had a more complex internal argument schema and were able to use the information given them in the Elaborated Goal condition to self-regulate their writing.

However, one might have expected the LD students to have reacted like the Grade 4 students, and this was not the case. 'Frankly we do not have a convincing explanation for the absence of interaction between goal and disability status' (p 700).

8. Despite the effects of instruction, the mean Grade 6 persuasiveness scores were not very high (elaborated goal: 2.5; general goal: 1.85), where 2 (out of 7) was,

'minimally developed; states a clear opinion and gives one or two reasons to support the opinion, but the reasons are not explained and supported in any coherent way' (p 700).

Moreover, only 54% of the Elaborated Goal Grade 6 students included alternative positions or rebuttals. Page-Voth and Graham (1999) showed that students responded to an alternative position when (a) taught to set goals to refute one and (b) learned a strategy to use.

Weight of evidence A (trustworthiness in relation to study questions)
High: This is to give the benefit of the doubt concerning sample representativeness, baseline writing differences and impact of ethnic membership (i.e. that the randomisation did indeed even out group differences).

Weight of evidence B (appropriateness of research design and analysis)
High: It was an RCT.

Weight of evidence C (relevance of focus of study to review)
High: The aims of the study are highly relevant to the review, as are the types of essay and the dimensions measured. The study does, however, represent the minimum possible intervention (closer to a testing situation, involving just the oral and written presentation of a goal-oriented prompt).

Weight of evidence D (overall weight of evidence)
High: This is to assume the sample was generalisable at least to the school district and that randomisation did reduce inter-group differences to minimal levels.

Graham S, Harris KR, Mason L (2005) Improving the writing performance, knowledge, and self-efficacy of struggling young writers: the effects of self-regulated strategy development. *Contemporary Educational Psychology* 30: 207-241

Country of study
USA

Age of learners
8-9: Grade 3

Type of study
Researcher-manipulated evaluation: randomised controlled trial

Aims of study
1. 'to examine the effectiveness of an instructional program designed to improve the performance of struggling young writers' (p 208) 'attending urban schools that serve minority and other children from mostly low income families' (p 234)

2. 'to examine if social support through peer assistance would enhance SRSD instructed students' performance' (p 210)

Summary of study design, including details of sample
This is an individual RCT with random allocation into one of three groups.
1. Struggling students isolated and allocated to groups
2. Pre-tested for skills at writing the two instructed and the two non-instructed genres, plus knowledge and sense of self-efficacy
3. Intervention: instruction in persuasive writing and stories
4. Post-tested for all four genres, plus knowledge and sense of self-efficacy
N = 73 randomly allocated, 72 completed the study

Nature of intervention
With the SRSD model, the students were explicitly and systematically taught strategies for accomplishing specific writing tasks. Students were also taught any information or skills needed to use these strategies. Students were also taught how to use self-regulation procedures including goal-setting, self-monitoring, self-instructions, and self-reinforcement to help them manage the target strategies and task of writing, as well as to obtain concrete and visible evidence of their progress.

With the SRSD plus peer support, the concept that students would act as partners to help each other apply the strategies they were learning to other situations and in other classes was introduced.

In the comparison condition, writing instruction was delivered to students by their regular teacher using the Writers' workshop model.

Data-collection instruments, including details of checks on reliability and validity
Assessments for each genre involved writing a paper in response to a writing prompt. Pupils were given a choice of two writing prompts.

The following were scored for each paper: composing time, length, total number of words, and compositional quality using a traditional holistic rating scale.

Persuasive essays were scored to see if they contained the basic elements that students in the two SRSD conditions were taught to include when planning and writing such papers.

Self-efficacy: instructor read item; student scored it. 5/10 items from Graham et al. self-efficacy scale.

Sample of double scoring for total word length (inter-rater reliability 0.99).

For each genre, examiners were provided with a representative paper for a low, middle and high quality score to serve as anchor points for their respective genre.

In addition, all papers were typed before scoring to minimise bias that might occur when examiners scored papers.

Compositional quality was scored by two former elementary school teachers who were unfamiliar with the design and purpose of the study. The quality score for a students' paper was the average score for the two raters - inter-rater reliability scores ranged between 0.72 and 0.93 (for persuasive it was 0.87)

Writing
1. Equivalence of prompts checked post hoc (no statistical differences, all ps >0.06).
2. Prompts from the pool were randomly paired to form sets which were counterbalanced across groups.
3. Sequence of genres tested was counterbalanced.
4. All students were individually tested.
5. Half of papers rescored by a graduate student re number of words: inter-rater reliability was 0.99.
6. All papers typed before scoring.

7. Holistic scoring was by two elementary school teachers; inter-rater reliability for persuasive writing and informational writing was 0.93 and 0.72.

8. Persuasive essays scored for elements by a graduate student, with half rescored by another; inter-rater reliability was 0.86.

Knowledge

9. Items read aloud to students as well as being on paper.

10. Responses to Q1-4 scored by two graduate students; inter-rater reliability for four categories was: production procedures (0.95), substantive processes involved in composing (0.91), motivation (0.71) and (innate) abilities (0.79).

11. Inter-rater reliability (again two graduate students) on Q5-6 was 0.93 and (productive procedures) and 0.78 (for elements).

Self-efficacy

Scores were factor analysed. Alpha coefficients for the two factors were 0.73 and 0.69.

The writing prompts were evaluated by a third-grade teacher and two third-grade children in terms of their suitability

Writing

1. Use of a choice of topics

2. Pre-screening tasks/topics by a third grade teacher and two third-grade children

3. Power not speed test format

Knowledge

4. Items read aloud by instructor

5. Don't know responses pursued: please think some more

Self-efficacy

6. Items read aloud by instructors

7. Scores factor analysed, with loading cut-off point at 0.4 (identical across factors/components)

Methods used to analyse data, including details of checks on reliability and validity

Non-parametric measures to test for differences across time and condition for the four genres

Kruskal-Wallis test to determine if there were statistically significant differences between the three conditions. Follow-up analyses were conducted, using the Mann-Whitney U procedure. Where scores were normally distributed, a one-way ANOVA was conducted.

Three-way ANOVA with repeated measures (self-efficiency data)

Checks on reliability and validity not applicable, although justification for the tests used is given

Summary of results

SRSD instructed students wrote qualitatively better stories and persuasive papers (the two instructed genres) than their peers in the comparison condition. The effects of SRSD instruction generalised to one of the uninstructed genres, as students in both conditions wrote informative papers that were qualitatively better than their counterparts in the comparison condition.

Composing time: Students in the two experimental groups spent more time composing their post-test persuasive essays than students in the comparison condition.

Length: There was a statistically significant difference between the three conditions in the length of persuasive essays.

Elements: There was a statistically significant treatment effect. Students on the two SRSD conditions included more basis elements than their counterparts in the comparison condition but there was no statistical difference between the SRSD conditions.

Quality: There was a statistically significant treatment effect. Students in both SRSD conditions wrote qualitatively better persuasive essays than comparison students.

Conclusions

'Teaching third grade struggling writers a general strategy for planning a composition, genre-specific strategies for story and persuasive writing, procedures for regulating the use of these strategies and the writing process as well as knowledge about the basic purpose and characteristics of the two genres had a powerful impact on the participating children's writing performance.' (p 234)

Weight of evidence A (trustworthiness in relation to study questions)

High: It is an RCT. The concern with baseline variables, checking the instructional procedures and using appropriate statistics mean that the results are convincing. On the other hand, the sample sizes of the groups are slightly small.

Weight of evidence B (appropriateness of research design and analysis)

Medium: In general, the trustworthiness is very high; the only missing feature for this review is that the maintenance probe only tested story writing, not persuasive writing.

Weight of evidence C (relevance of focus of study to review)

Medium: Due to small sample size - all other factors highly appropriate. The measures that were used have a high weight of evidence, with the possible exception of the self-efficacy measures.

Weight of evidence D (overall weight of evidence)

Medium to high: Medium simply due to the limitations imposed by the lack of a persuasive writing maintenance probe, the indeterminacy about the implications of the self-efficacy data, and possibly the fact of the comparison group being the only group to be taught and tested by their regular teachers. Otherwise, the study is of high trustworthiness

Hammann LA, Stevens RJ (2003) Instructional approaches to improving students' writing of compare-contrast essays: an experimental study. *Journal of Literacy Research* 35: 731-756

Country of study

USA

Age of learners

13-14: 8th graders at a large eastern middle school

Type of study

Researcher-manipulated evaluation: controlled trial

Aims of study

The goal of this study was to examine instruction designed to support students' knowledge acquisition and text structure organisation as related to compare-contrast writing' (p. 732).

Summary of study design, including details of sample

This is a controlled cluster trial of five intact classes assigned to one of three interventions or control (two classes). The authors claim that it was not possible to assign participants to treatment groups randomly. Therefore, following quasi-experimental procedures, they randomly assigned intact classes to treatments, using a 2 x 2 pre-test/post-test design and analysis of covariance to control for pre-test differences among groups (p 735). The participants were students in five

intact eighth-grade classes. Although the number of students who participated in the study is unclear, 63 students' essays were used in the data analysis.

Nature of intervention
The treatments were devised and taught by one of the authors of the article. The teaching was incorporated as part of the students' language arts class. There were three treatment groups: Summarization Skills Treatment, Text Structure Treatment or Combined Treatment. In each of the teaching sessions, the students were first introduced to the particular topic which was to be taught. Depending on which group the teacher was working with, the teaching sessions included a particular focus on teacher modelling of either summarisation skills or text structure, or a combination of both. Issues of continuity and progression were stressed in each of the teaching sessions. Students received written feedback on the strengths and weaknesses revealed in their written work.

For pre and post-test measures, students were asked to read two information texts (written by one of the authors of the article) as a prompt for writing a compare-contrast composition using the information located in the texts. Each set of two source texts consisted of two topics: 'desert set' or 'ruins set'. Students' writing was analysed according to two scoring rubrics: content measure and structure measure.

Only students who returned informed consent forms signed by their parents had their measures included in the study. There were 36 females and 27 males whose essays were used in the data analysis.

Data-collection instruments, including details of checks on reliability and validity
The same researcher administered the intervention for each of the control groups. The instructions given to the students for the pre and post-test writing exercises were written down, so that each teacher would say the same thing. Every student wrote a compare and contrast piece on each of the two topics - either as a pre- or a post-test exercise. The pre- and post-test essays were written under timed conditions (one class period, presumably of 45 minutes). The assumption is that the data was collected at the end of this writing period by the teachers who read the directions to the students and supervised the writing period.

One of the researchers wrote the reading material used for the pre- and post-test exercises (and all but two of the pieces used for practice). This material was checked for suitability with the classroom teachers. It was also checked for readability and for Prior Knowledge.

The researcher discussed the idea unit (see Content measure) with five graduate students who then independently identified idea units from several practice texts, resolving differences by discussion and consensus. They then found idea units in the pre-test and post-test texts and computed the average number for each (pp 741-742). Five other graduate students rated the level of importance of each of the units. Average values for each idea unit were calculated. 'This provided a list of ideas in each text by their level of importance to use in determining the number of higher-level ideas students used in their writing' (p 742).

The rationale for the compare and contrast essay writing exercise was informed by a substantial body of research literature into the conceptual challenges posed by that medium (see, for example, Englert and Thomas, 1987; Sitko, 1998; Spivey, 1990, 1991) (p 739). The Prior Knowledge test was also informed by research literature, though the authors point out that research studies have not provided a generally accepted and reliable assessment method (p 740).

Methods used to analyse data, including details of checks on reliability and validity
Two scoring rubrics were constructed, one for Content measure, (the number of important ideas used from source texts) and the other for Structure measure (compare-contrast text organisation of the paper) (p 741). For the Content measure, the number of idea units per text was identified and ranked in order of importance on a scale of 3 (important) to 1 (not very important). 'The Structure rubric had seven subscales: four to measure primary traits of compare-contrast organisation, one to identify overall compare-contrast organisation, and two to provide quality measures of writing (p 742).

A series of 2 x 2 ANOVAs on all pre-test essays, collapsed across all source text topics, was conducted for each of the measures. A series of 2 x 2 ANCOVAs on all post-measure essays, again collapsed across all source text topics, was conducted to assess outcome measures for content and structure. The covariate for each measure was the pre-test score for that particular measure (p 743).

Graduate students were trained to score the students' essays. Pre-measure and post-measure essays were randomly assigned to rater pairs, with master codes on each, so that raters could not identify either treatment group or pre/post-test conditions. Each pair of raters scored all assigned essays on both the Content and Structure rubrics. Final scores were based on the average of the two raters' scores. No rater saw another's rubrics and scores (pp 742-743).

Summary of results
Students receiving Text Structure instruction had significantly higher scores on compare-contrast Structure but lower ones on Content than did those in the other groups. Students receiving Summarization Skills instruction had significantly lower scores on Structure than did students in the other groups (p 731). Although the pre and post-test texts were tested for Prior Knowledge, students still performed better on the 'ruins' assignment than on the 'deserts' topic.

There was a statistically significant main effect for Text Structure treatment on measures of Content in students' writing on the 'deserts' task. Students receiving Text Structure instruction had lower means on the Content measure than students who had not received it. For the Structure measure, there were no significant effects for Summarization Skills treatment, Text Structure treatment, or Combined (p 744).

For the 'ruins' task, there were no statistically significant effects on the Content measure. However, the means for the Text Structure group was higher than the other three means. There were statistically significant main effects for both treatments on measures of Structure. There was no significant interaction between the two treatments (p 744).

Conclusions
The authors conclude that the results of the study provide further support for explicit text structure instruction for writing already highlighted by previous research (p 746). They suggest that content is an important factor to consider in the teaching of writing. Simple instructional techniques focusing on compare-contrast or summarisation or text structure are not enough. Task and strategy instruction must be matched. Further research is required into the complex nature of transfer. Learners may not perceive a need to utilise a strategy if they think they already understand the material.

Weight of evidence A (trustworthiness in relation to study questions)
Low: The intervention was very short and no follow-up work was conducted. Very little contextual information is provided about the sample. The study was not randomised. The authors do not consider issues of selection bias, nor do they mention the issues of sample size in a cluster trial and the clustered nature of the data in the analyses.

Weight of evidence B (appropriateness of research design and analysis)
Medium: The research design is appropriate but there could have been randomisation plus a follow-up to the post-test. Although detailed, thorough and thoughtful in its application, the study was very short. The authors do not consider the possibility that the results could be confounded by teacher effects and/or selection bias. The issue of sample size in a cluster trial and the clustered nature of the data in the analyses are not mentioned.

Weight of evidence C (relevance of focus of study to review)
Medium: The conceptual focus of the study and the measures used certainly address important features of content and structure which are germane to the review question. However, the exclusion of children classed as Learning Support and ESL means that the study does not explore the whole ability range. Little information is provided about the 63 students whose essays were used for data analysis. The issue of Prior Knowledge, though addressed by the authors, is not resolved satisfactorily (p 752).

Weight of evidence D (overall weight of evidence)
Low to medium: The authors themselves acknowledge the limitations of the study, the most notable of which are the timescale and the lack of balance in the pre and post-test texts. Nevertheless, this is thoughtful work which is appropriately modest in its claims and which is grounded in the research literature. It gives a sound basis for further research into the links between ideas (content) and structure in argumentative writing for 8th-graders.

Hidi S, Berndorff D, Ainley M (2002) Children's argument writing, interest and self-efficacy: an intervention study. *Learning and Instruction* 12: 429-446

Country of study
Canada

Age of learners
Mainly 11-12

Type of study
Researcher-manipulated evaluation: controlled trial

Aims of study
The purpose of this study was to examine how a combination of motivational and instructional variables can best be utilised in an intervention programme to improve students' emotional and cognitive experiences during argument writing (p 429).

Summary of study design, including details of sample
This is a controlled clustered trial with two groups, using pre-test, two forms of interventions and a post-test, plus questionnaire for all students. Procedures for allocation are not described. The sample consisted of 180 students from four (presumably intact and non-randomised) Grade Six classes and four junior intermediate classes. Participants were divided into two intervention groups of different size. It is not clear how many pupils were in each group. All eight classes received the basic intervention programme (IP), but six of the eight also received the additional motivational component (IP + M). Pre- and post-testing took the form of written assignments on essay topics selected by the researchers. There was no control group.

Nature of intervention
An IP intervention programme was administered over eight weeks. This consisted of classroom discussion of sample arguments; lectures and notes on the characteristics of argument writing; classroom exercises in looking at a problem or issue from another point of view; and a teacher-assisted collaborative argument-writing exercise.

For six of the eight classes, an additional motivational component was incorporated into the basic IP instructional programme. It consisted of students working in small groups of five to write arguments on interesting topics and exchange by fax their writing with students at another school. Student groups were then paired to exchange arguments, provide feedback, make suggestions and request clarification within these larger groups. This information was used by the student groups to revise their original drafts.

Students' written work was analysed at pre and post-test level according to a holistic rating score ranging from one to six. Criteria included maturity of style, ability to reason persuasively, observance of the conventions of written English. Their work was also scored for position or thesis statement, supportive statements and evidence, treatment of the opposing view and use of a conclusion.

Data-collection instruments, including details of checks on reliability and validity
Students wrote argumentative essays on topics approved by the teachers and informed by research literature on children's argument writing (for example, Knudson, 1992). The titles were chosen by the researchers according to the following criteria: interest to students, student knowledge and similarity of content (p 435). The pre-test title was: 'Should students go to school six (6) days a week?' and the post-test title was: 'Should there be less school holidays?' In addition, the students completed a questionnaire pre and post-intervention. They were invited to give a written answer to the question: 'Do you like writing?' They were also asked to rate on a five-point Likert-type scale how much they liked and how good they thought they were at doing eleven text types that they were likely to be involved in at school. (p 437)

The study design and framework was informed by the work of Boscolo. The questionnaire was modified from a design constructed by Boscolo and Cisotto (1999) and from the Writing Self-Efficacy Instrument for adults produced by Shell et al. (1989).

Methods used to analyse data, including details of checks on reliability and validity
To evaluate all pre- and post-test arguments, a holistic rating scale was developed with scores ranging from one to six (p 437). Top marks of six were awarded to students who wrote in a mature style, reasoned persuasively, observed the conventions of written English, and included the following four components: position or thesis statement, supportive statements and evidence, treatment of the opposing view, and conclusion. Lower scores reflected writing that was missing one or more

of the four argument components and/or was less well written (p 437).

Student responses to the questionnaire question 'Are you interested in writing?' were coded into three categories: Yes, No, or Ambivalent (p 437). To measure specific liking and self-efficacy, students rated 11 text types on a five-point Likert-type scale.

The writing scores were examined for gender and school difference. A three-factor (gender, school, and intervention type) repeated measures ANOVA revealed a significant main effect of pre-post test. The same type of analysis was repeated for the word count scores. The data were screened statistically for outliers. Eleven were eliminated from the data analysis, in order to prevent a small number of extreme cases from distorting the group statistics. (p 438)

A graduate assistant and one member of the research team used a sample of arguments written on the pre-test topic by students who did not participate in the project to refine the scale and, subsequently, to practise scoring arguments using the rating scale (p 437). Inter-rater reliability was high.

Validity is addressed only in so far as the focus of the analyses is based upon issues identified by earlier researchers such as Boscolo or McCann.

Summary of results

The quality and the quantity of the arguments students produced significantly improved for both boys and girls from pre- to post-test, although the improvement was not the same across all the schools. The additional motivational component (IP+M) of collaborative writing significantly improved boys' performance more than girls' (pp 441-442). Collaborative group work and exchanging information between schools did not affect all students equally (p 441).

From the questionnaire, it was found that 66% of the students said they were interested in writing pre-test, and 72% post-test. Affect and efficacy were closely related for each of the writing types (genres) that were asked about, and there was little difference between boys and girls in this respect, with girls just marginally showing higher affect and efficacy scores.

In terms of students' attitudes to writing and preference for particular text types, the authors report 'remarkable' consistency among the means and the rank order of means for the various genres across affect and efficacy ratings. While not significantly altering levels of affect and efficacy, the intervention may have strengthened the association between them (p 440).

Conclusions

General interest in writing, enjoying writing in several genres and feeling efficacious about such writing are closely related. Providing students with challenging tasks, allowing them to use prior knowledge, requiring them to use self-management and self-regulatory learning strategies, and providing opportunities for collaboration with peers - all play a part in improving students' written arguments. Only the fourth factor did not affect participants in this study equally (being more efficacious for boys). The authors suggest that, when boys are not interested in an activity, they are more reluctant to be compliant and perform than similarly uninterested girls. Once their interest is aroused, boys' performance appears to improve significantly (p 442).

Further work needs to be done into determining how interest, enjoyment and 'feeling efficacious' about writing can best be utilized to improve students' performance (p 442) and into the reasons for the variations in performance between schools noted in this study.

Weight of evidence A (trustworthiness in relation to study questions)

Low: There is insufficient information about the nature of the interventions, little exploration of the content of the students' essays and no follow-up to the post-test in what is actually quite a short intervention.

Weight of evidence B (appropriateness of research design and analysis)

Medium: This is a detailed and carefully constructed research design but, given that the aim is to test how interventions affect motivation for, and the quantity and quality of writing arguments, a better design would have been a randomised trial with a control group, including a delayed post-test.

Weight of evidence C (relevance of focus of study to review)

Medium to low: The study focus – self-efficacy, interest, argumentative strategies and issues relating to purpose, context and audience - is germane to the research question. The age group is also appropriate. However, an undisclosed number of the students in the sample spoke English as a second language. It is not clear whether the entire ability range was included in the cohort, nor is it clear how many males and how many females were involved.

Weight of evidence D (overall weight of evidence)

Low to medium: Lack of context (particularly in relation to the gender element of the study), time limitations, lack of detailed reporting of the students' writing. The shortcomings in the research design makes for a rather heavyweight analytical framework on the basis of a poorly conceived sample and thin material (not described in detail; students' work not seen). So the study is unsatisfying. Its results are not surprising, but neither are they very valid or reliable.

Knudson RE (1991) Effects of instructional strategies, grade, and sex on students' persuasive writing. *Journal of Experimental Education* 59: 141-152

Country of study
USA

Age of learners
9-13: Students from grades 4, 6 and 8

Type of study
Researcher-manipulated evaluation: randomised controlled trial

Aims of study
To identify the effects of instruction, grade level, and sex on students' persuasive writing.

Summary of study design, including details of sample

This was a randomised controlled trial in which there were four 'treatments': instruction with model pieces of writing; scales and questions designed to guide students' writing; both models and scales and questions; and a control group with no explicit instruction in persuasive writing.

The study involved 159 students from across the primary/secondary transition divide. It was conducted over a relatively short time span and, although there was follow-up work, this took place only once, two weeks after the intervention finished. The intervention materials were designed to teach persuasive writing using

four instructional strategies involving both reading and writing opportunities (details are provided in Appendix A). The control group received no explicit instruction in persuasive writing and was invited, instead, to practise writing narrative.

Nature of intervention

There were four treatments, as described above:

(1) Instruction with model pieces of writing

(2) Scales and questions designed to guide students' writing and revision

(3) Both model pieces of writing and scales and questions

(4) The control group was shown a picture every day and asked to write a story about what was happening. They were not given explicit instruction in persuasive writing.

Students from two schools were involved in the study: 55 fourth-grade and 55 sixth- grade students who attended an elementary school and 49 eighth-grade students who attended three English classes in a junior high school located in the same urban area of south-ern California. Two sets of materials were designed for the intervention. One set was used with students from grades 4 and 6, and the other with students from Grade 8. The materials were designed to teach persuasive writing, using four instructional strategies. Students were randomly assigned by grade to treatment groups. The intervention lasted for 14 days at 20 minutes per day. Writing samples were collected from the students immediately after the intervention and then again two weeks later.

Data-collection instruments, including details of checks on reliability and validity

Details are provided as Appendix B of the original article. Each writing prompt required a piece of persuasive writ-ing as a response. In two of the three examples given in Appendix B, the medium for the piece of writing (a letter) was also stipulated. An example is: 'Write a letter to the school principal to convince him/her that there should be more school holidays' (p 151).

The reading material used in the intervention was checked for readability using the Fry Readability Formula (Fry, 1968). It is not clear whether the writing prompts were included in this check. The writing prompts were, however, 'carefully written so the audience and purpose were clearly expressed in each instance' (p 144).

The authors modified Hillocks' (1986) meta-analysis of instructional strategies in writing for use in the interven-tion.

Methods used to analyse data, including details of checks on reliability and validity

'The writing samples were assessed with a modified version of Diederich's (1974) scale' (p 144). The authors used six of Diederich's original eight criteria: clarity, organisation, coherence, punctuation, spelling, and word choice. Scores were allocated on a four-point scale, ranging from 1 ('not acceptable') to 4 ('very accept-able) (p 151). Details of the scoring guide are provided in Appendix C.

Mean and standard deviation analysis for 'Content' and 'Form' in the persuasive writing samples. There was a Scheffe test to compare grade and content.

Two raters scored the material. 'If the two raters dif-fered by no more than 1 point, the scores were summed ... If the scores differed by more than 1 point, the scores were settled by discussion between the raters. The origi-nal scores were used to calculate inter-rater reliability' (pp 144-145).

Cronbach Alpha tests were run to check the six scores given for each essay during the analytic assessment for internal consistency.

Validity is addressed only in so far as they draw upon the work of Diederich and explain how and why they modified the original eight-point scale.

Summary of results

For both writing samples, four of the scores appear to be moderately highly inter-correlated: clarity, organisation, coherence and word choice.

There was a significant main effect for grade for the content score in that 'grade 8 wrote better than students in grades 6 and 4, and students in grade 6 wrote bet-ter than students in grade 4 for Writing Sample 1. 'For Writing sample 2, however, students in grade 8 wrote significantly better than students in grades 6 and 4, but there were no significant differences between the mean scores of students in grades 6 and 4.' (p 147)

For the form score, 'there was a significant main effect for grade' (ibid.).

Furthermore, 'although girls wrote significantly bet-ter than boys at the conclusion of treatment, after two weeks boys and girls received similar scores on their essays' (ibid), i.e. girls dropped off in performance two weeks after the treatment.

Conclusions

The authors conclude that, to a certain extent, their study confirms the work of other researchers that 'chil-dren have difficulty with persuasive writing' (p 149). This study also confirms earlier research by indicating that older children write better than younger ones. The authors add: 'The results of this experiment, however, indicate not only that older children write better than younger ones, but that eight-grade students perform equally well immediately after treatment and 2 weeks after treatment, whereas fourth- and sixth-grade chil-dren decrease in their ability to respond to persuasive written tasks' (p 149).

However, the authors acknowledge that there are important findings upon which they can cast little light. For example, they concede that 'there is little to explain why girls' scores would drop so dramatically when treat-ment was discontinued' (p 148). They also attribute the fact that 'results were mixed and inconclusive' (p 149) to a number of limitations to their study design.

It is unclear whether this loss in writing skills is due to the lack of cumulative practice and instruction in persuasive writing, or whether eighth-grade students are able to absorb the writing strategies into their repertoire of writing abilities because they exhibit the more devel-oped logical thinking usually regarded as a prerequisite for effective persuasion/argumentation' (p 149).

Weight of evidence A (trustworthiness in relation to study questions)

Medium: The study is high on reliability but only medium on validity. It is not entirely possible, therefore, to trust the results. We do not actually see much of the interven-tion itself.

The fact that this study is a randomised control trial gives it weight. The data is subjected to some rigor-ous statistical analysis. However, the intervention is brief and the maintenance probe was administered very shortly after the intervention ended. Little evidence is provided to support the validity of the intervention or data analysis. The rationale for the four point essay-rat-ing scale lacks detail. Why, for example, does a spelling error have to occur several times before it is counted as one error? Who is to judge which words count as 'vulgar' or 'trite' in the (very subjective) rating scale relating to 'word choice'? The authors concede that the research

design did not provide access to important information as effectively as it might have done. There is very little contextual information about the sample distribution and very little information about the precise nature of the intervention.

Weight of evidence B (appropriateness of research design and analysis)
Medium: The research design is highly appropriate, in that it adopts a randomised approach at the individual level to determine whether there is any effect as a result of the interventions. There are some doubts about the validity of the study.

However, it is not clear whether the sample includes the whole ability range of students, n. or is it clear whether English is a first language for all the students concerned. The absence of baseline assessment in writing (evidence is only forthcoming for reading comprehension) or pre-test samples of students' persuasive writing makes it a little difficult to judge the effect of the intervention - which is itself very short and only subjected to a maintenance probe two weeks post-test. Not much evidence is offered to support the claim that the data collection and analyses procedures are valid.

Weight of evidence C (relevance of focus of study to review)
Medium: More contextual information about the sample is needed for a higher rating to be awarded. As mentioned above, the lack of baseline and pre-test assessments is a drawback. Similarly, as the authors themselves concede, the focus upon 'global writing ability' (p 149) means that important information about the form scores on the writing samples is not forthcoming. 'If the primary criterion for 'good' persuasive writing had been that the response have the form of an argument, statistically significant treatment effects might have been found' (p 149). 'It is possible that treatment results might have become significant if the length of treatment is increased' (p 149). There is little qualitative information about the content of the writing samples. The four point essay-scoring scale seems vague and subjective.

Weight of evidence D (overall weight of evidence)
Medium: Medium in terms of the validity of the design and trustworthiness of the conclusions. The author appears to fall back on Piagetian possible explanations when her intervention is seen to actually reduce the performance of students over the experimental period, and when the only conclusion appears to be that older children write better than younger ones (not very surprising!). This represents a conservative approach to research.

The authors concede that some of their results were 'mixed and inconclusive' (p 149). They have little to say about the fact that the girls' scores dropped 'dramatically' when treatment was discontinued (p 148). It is not clear why the conclusion that older children write better than younger ones is 'particularly interesting' (p 148). The authors' conclusions as to why children have difficulty with persuasive writing seem a little rigid and also theoretically questionable. There are issues about timing, baseline measurements, context and validity.

1. Knudson RE (1994) An analysis of persuasive discourse: learning how to take a stand. *Discourse Processes* 18: 211-230

2. Knudson RE (1992) An analysis of persuasive discourse: Learning how to take a stand. Paper presented at the Annual Meeting of the National Reading Conference, San Antonio, TA

Country of study
USA

Age of learners
8-11: grades 3 and 5

Type of study
Researcher-manipulated evaluation: randomised controlled trial

Aims of study
To determine the effects of instruction on students' persuasive writing at two grade levels (3rd and 5th) and to determine the categories and types of written persuasion used by students at four grade levels (3rd, 5th, 10th, and 12th).

Summary of study design, including details of sample
This was a randomised control test with two aims: first, to determine the effect of instruction (written or written and oral) on persuasive writing at two grade levels (3rd and 5th); and second, to determine the categories or types of written persuasion used by students selected from across the primary/secondary divide (Grades 3, 5, 10 and 12).

The authors assert that 'This study employed a repeated measure, 2 (Oral Interaction) x 4 (Treatment) x 2 (Grade) x 3 (Time of Measurement) design' (p 215). However, this does not entirely apply to the 'Classification of Persuasive Statements' section of the study (p 217). The latter involved students 'who responded to one, not three, [writing] prompts' (p 217). (The 10th and 12th graders had, however, undertaken a similar treatment - minus the oral component - in an earlier study).

Explicit for the 'classification of persuasive statements' section of the study: papers of 313 students.

139 students 'were present for all three writing prompts and at least 10 of the 14 days of instruction' (p 214).

Nature of intervention
The intervention consisted of four elements:
1. Presentation of model
2. Presentation of scales/questions/criteria
3. (1) and (2) combined
4. Free writing plus, oral interaction for some classes.

The study involved a randomised control test involving 139 mixed sex students in eight classes from Grade 3 to Grade 5 in the same elementary school in California. Variables under consideration were argumentative strategies, gender, treatment, grade and time. The four levels of treatment were administered over a 14-day period.

One classroom at each grade level was randomly selected to receive only textual instruction for the 14-day intervention period. These two text-only classes served as a control group. The remaining classrooms received nine days of written/pictorial instruction. They also received an 'oral interaction component' (p 216) for five days.

All students completed writing samples pre-test, at the end of the intervention and again two weeks later. Each of the prompts involved writing a persuasive letter to the school principal regarding an aspect of school life.

Data-collection instruments, including details of checks on reliability and validity

'The test samples were collected under uniform conditions' (p 216).

'Before the study started (Writing prompt 1), at the end of the study (Writing prompt 2), and again 2 weeks after the completion of the study (Writing prompt 3), writing samples were collected from all the students' (p 216)

The writing prompts are set out in detail in the paper. They basically consist of simple invitations to write argumentative/persuasive essays in the form of 'Write a letter to the school principal to convince him/her of x...'

Prompts 'were carefully written so the audience and purpose were clearly expressed in each instance' (p 216). 'Treatment variations did not apply to the production of the samples' (p 216). 'Treatment remained constant within each treatment group' (p 215).

'The four methods of instruction were selected, in part, on Hillocks' (1984, 1986) meta-analysis of the effectiveness of instructional strategies in teaching writing' (p 213). The decision to include two text-only classrooms as a control was informed by previous studies conducted by the author (Knudson, 1993, 1992, 1991).

Methods used to analyse data, including details of checks on reliability and validity

Student papers were scored holistically, taking account of the purpose for writing, its audience, and the degree to which the task was addressed. A six-point scoring guide was used, with 1 being the lowest and 6 the highest. A paper scoring full marks would have been deemed to have addressed the topic, stated and elaborated arguments and exhibited logical thought (see Appendix B for details).

In the case of the 313 essays written by students from grades 3, 5, 10 and 12, argumentative strategies (see Appendix C) were subdivided into the following five categories: norm invocation, positive sanction, negative sanction, request and assertion (see Table 1 on page 219).

'The data were analyzed with Statistical Analysis System (SAS, 1985) using the general linear model procedure ... Because this is a nonorthogonal repeated measures analysis of variance, sum of squares III is used' (p 218).

Two raters evaluated the writing samples independently, scoring each essay. Inter-rater reliability was high. For the 'classification of persuasive statements' section of the study, whenever there was a difference in the strategies listed, one of the raters listed an additional strategy in the paper. All differences were settled by discussion and the discussions helped the raters re-anchor the categories. Because 'reason' was the most frequently used category in requesting a change in school rules, the raters were asked to analyse the responses in this category. Inter-rater reliability regarding categorisation of the students' choice of reasons was high. A further check was carried out by the researcher.

The two essay readers were 'trained in the classification system of Weiss and Sachs (1991) ... which was based on a modified version of Falbo's (1977) 16 strategies' (p 217). Falbo's categories were 'collapsed into five types, modelled after Wood, Weinstein, and Parker (1967) and Weiss and Sachs (1991)' (p 218).

Summary of results

In terms of classification of persuasive statements, analyses revealed that 'there were no significant main effects for gender', although there were 'significant main effects for grade'. There was 'no significant interaction effect for Gender X Grade'. The 'number of strategies employed increased significantly with grade' (p 220). There was a 'significant difference in the use of two categories for the dominant message by grade: Category 5 (Compromise) and Category 24 (Simple Statement)' (p 220).

'Analyses on the use of types resulted in similar findings. Type 3, Negative Sanction, was used more by students in Grade 5 ... than by students in Grade 3 ... Type 4, Request, was used significantly more by students in Grade 3 ... than by students in Grades 10 ... or 12' (p 220).

The use of reason 'was a frequently used dominant strategy.' There was 'a significant main effect for gender ... but no significant main effect for grade ... There was no significant interaction effect for Gender X Grade ... girls used significantly more reasons than boys' (p 221).

The results for the analyses of reasons used are depicted in Table 3 (p 222). 'The students appealed to the welfare or good of specific people or groups of people under the general categories of Safety, Convenience, or Pleasure when asking for a change in a school rule. Convenience was the most frequently used category, and Other Students/Friends were cited most often as the ones who would benefit from the change in a school rule' (p 221).

'There were no significant effects for treatment or for oral interaction for the instructional intervention' (p 221).

Conclusions

The author concludes that little is known about what makes a good persuasive argument or about how to teach effective argumentation. 'What we do know ... is that third-grade students have simple arguments with a larger percentage of requests than at other grade levels. Fifth-grade students' arguments are more complex than third-graders' arguments, but less complex than 10th- and 12th-graders' arguments. Fifth-graders 'use more negative sanctions than students at any other grade level. Tenth- and 12th-graders' arguments are very similar in terms of complexity and kinds of statements used ... There appears to be a definite gradual addition of Compromise by grade to students' arguments'. 'Very few students at any grade level wrote about the opposing position, or why the school rule should not be changed'. She does note that the lack of statement of opposing arguments by the students appears to hold them back. 'What is obvious here is that there is a growing sophistication by grade of what works in making a written argument to an adult in a position of authority (pp 222-223).

Weight of evidence A (trustworthiness in relation to study questions)

Medium: The results can be trusted to an extent, in that, as in many American studies of this kind and this period, the reliability factor is high. But reliability is won at the expense of validity: there is little given in way of the context of the study, the quality of the intervention or indeed of the invitations to write for the pre-test and post-tests.

The intervention groups participated in a randomised control test. However, the lack of information about the study involving the Grade 10 and Grade 12 students is a drawback. More contextual information about who chose what essay topic and about who missed which of the carefully demarcated intervention days would have helped. It would have been useful, too, to see how the raters actually analysed some samples of students' writing, in order to get a sense of how the six-point scoring guide worked in practice. Information about the oral

intervention is a little vague: it is not till the 'discussion' section of the paper that the author comments: 'It should be noted that the oral interaction was of a recitation/discussion format' (p 221). More contextual information on this point might help to explain the 'very disappointing' results for oral interaction (p. 221). The intervention was brief and the maintenance probe conducted only two weeks after treatment.

Weight of evidence B (appropriateness of research design and analysis)
Medium: As a randomised controlled trial, this study ought to have been able to indicate some clear results about the effectiveness of the interventions. However, not enough information is available regarding the work done with the Grade 10 and 12 students.

Weight of evidence C (relevance of focus of study to review)
Medium to low: The rationale for the inclusion of the oral component or the conceptual links between oral and written persuasive language are not clearly explored. Nor is there much information about the content of the oral intervention or the about why particular strategies were chosen over others. As year 10 and year 12 students were included in the study, the study as a whole is not highly relevant to the present review.

Weight of evidence D (overall weight of evidence)
Medium: This is an unimaginative study which nevertheless attempts to come to a clear understanding of the effect of certain (not very clearly specified) interventions on the writing quality of third and fifth graders.

Despite the use of randomised control and experimental groups and the detailed statistical analysis of the data, the actual findings of this study seem somewhat unremarkable. That students' grasp of argumentative strategies improves with age and maturity does not come as a great surprise. A lot of work seems to have been undertaken in order to arrive at the conclusion that 'little is really known about what makes a good persuasive argument' and 'even less is known about how to teach effective argumentation' (p 222).

Reznitskaya A, Anderson R (2001) Influence of oral discussion on written argument. *Discourse Processes* 32: 155-175

Country of study
USA

Age of learners
9-11: grades 4 and 5

Type of study
Researcher-manipulated evaluation: controlled trial

Aims of study
'This study aims to provide evidence about the effects of discussions in which children engage in oral argumentation on the reasoning that the children then exhibit in persuasive essays' (p 157). It 'examines whether oral discussions can help students acquire 'portable' knowledge of argumentation' (p 159).

Summary of study design, including details of sample
This was a clustered controlled (non-randomised) trial, using existing classes as clusters. There were four classes involved in four schools: two experimental and two control classes (115 students in all).

It is a quasi-experimental research design and has therefore the limitations associated with such studies. Intact, non-randomised classes from four public schools were involved in the study as either control or experimental groups. An attempt was made to match control and experimental groups in terms of ethnicity, socioeconomic composition and age. Classes whose teachers had previously introduced CR into their instruction procedures were not allowed to act as control groups. Following a baseline assessment of vocabulary, the control groups received their regular language arts instruction for a five-week period while the experimental groups were instructed in CR and also engaged in discussions with other participating students via web forums. At the end of the intervention, all the participating students were given 40 minutes (in class, presumably) to write a persuasive essay which explored a moral dilemma. The essays were scored to see if there had been any translation of the oral reasoning skills taught and explored during the CR intervention into the written work of the students.

Nature of intervention
The intervention consisted of discussion of controversial issues; coaching by teachers in formal argument devices; and web forums.

Students and teachers from four public schools in central Illinois participated in the study. Two same-grade classrooms were selected in each of two schools, one of which participated in CR discussions and one of which received only regular language arts instruction. Students in the CR classrooms met twice weekly in small groups of between 6 and 8 participants to discuss controversial issues for between 15 and 20 minutes. Teachers coached these students in the specific formal argument devices that promote the development of reasoned. The CR students also engaged in 15-minute CR discussions with the other participating classrooms via Web forums. At the end of the five-week intervention, students from CR and contrast classrooms wrote a persuasive essay based on a moral dilemma. The essays were coded to measure students' ability to consider a variety of relevant arguments, counter-arguments and rebuttals as well as to use evidence and to employ certain formal argument devices.

Data-collection instruments, including details of checks on reliability and validity
For the vocabulary test, students had to answer questions based on a list of 88 words and nonwords. For the written task, they had to respond to a scenario outlining a moral dilemma involving two boys called 'Thomas' and 'Jack'. Their task was to 'write an essay answering the question of whether Jack should tell on Thomas' (for cheating in a competition) (p 161).

They argue that 'tests of this type [the baseline vocabulary test] have been shown to be reliable and valid measures of vocabulary knowledge (Anderson and Freebody, 1983)' (p 160). No information regarding the reliability of the essay writing procedure is provided.

The essay-scoring criteria seem to have been informed by the relatively large body of literature into the effects of CR instruction described on page 159.

Methods used to analyse data, including details of checks on reliability and validity
'Essays were coded in five steps' (p 162):
(1) Essays divided into ideas units that represent the distinct parts of a claim.
(2) Each ideas unit is classified according to whether it can be 'coded' or 'not coded'. Coded units are those which are judged to be clear and relevant to the main question and which contain formal argument devices.

Non coded units are those which are judged to be unclear, supplementary or irrelevant.

(3) Coded units are then assigned to one of six categories relating to argumentative procedures: position, argument, counterargument, rebuttal, form and repeat.

(4) Units categorised as argument, counterargument and rebuttal are further classed as 'textual information' (p 162).

(5) Units categorised as 'form' are further subdivided as either 'explicit reference' or 'perspective' (p 163).

In addition, a sample of students' essays were reviewed qualitatively (see pages 168-171).

'All essays were scored by one rater blind to whether the essay was written by a student from a CR classroom or a contrast classroom. Scoring was checked at least twice to ensure agreement with the coding system. Twenty-nine essays were randomly selected to check inter-rater reliability. These essays were scored by a second rater who had been trained in applying the coding system.' (p 163) Correlation for coded category between the two raters: 0.96. Correlation for the irrelevant category: 0.99.

There are references to previous research into CR.

Summary of results
'In this study, students who participated in CR discussions wrote essays that contained a significantly greater number of arguments, counterarguments, rebuttals, uses of formal argument devices, and references to text information than the essays of similar students who did not experience CR' (p 171). The essays written by CR students also contained more words.

Conclusions
The authors conclude that 'reasoning skills acquired in discussion transferred to a different context, from oral collaborative discussions to the individual task of persuasive writing. This finding is consistent with the more general thesis that participation in oral argumentation promotes individual reasoning' (p 171). They suggest that because an argument schema is abstract, 'it enables transfer between different contexts and communicative modes' (ibid.) and also suggest that children in 4th and 5th grades can benefit from such approaches.

They also add that collaborative discussion 'appears to be an effective training ground for the development and internalization of generalized knowledge of argumentation' (p 173).

'With suitable experience, even children as young as the participants in this study can make considerable progress in the elaboration of a useful schema. Our theory is that children generalize elements common in CR discussions. These common elements include formulating an opinion, supporting it with reasons, presenting and responding to counterarguments, and using certain rhetorical forms' (p 172).

However, the authors issue a caveat: 'the findings of this study should be interpreted with the caution due in any quasi-experimental research' (p 173) They list a number of questions which the study failed to answer. For example, why was there no significant difference between the results of the experimental and control groups in School B? Can a writing exercise provide students with a full opportunity to reveal the reasoning skills they have acquired orally, in group situations? Should an 'individual oral argumentation task' have been included in the study? They conclude that further research is needed.

Weight of evidence A (trustworthiness in relation to study questions)
Medium: The research design is that of a quasi-experimental study with clustered groups. It is not entirely clear why the study adopted a vocabulary test as a pre-test measure, nor is the validity of the data collection or analysis made explicit. Although the data analysis is highly reliable, one wonders why the study was done like this.

The authors admit themselves that their results need to be taken with caution. The idea of measuring reasoning skills exclusively through a written outcome is, as the authors themselves concede, questionable. It is not clear why measurements of vocabulary were included, pre- and post-test, in a study concerning argument. Here again, the relationship between 'reasoning' and 'argument' is not clearly teased out. The rationale for the choice of essay scoring rubric is not explained properly. The supposedly 'detailed analysis' of selected students' essays (pp 168-171) seems rather subjective. Contextual information is lacking. There does not seem to have been any follow-up work.

Weight of evidence B (appropriateness of research design and analysis)
Medium: The authors admit that a quasi experimental study of this kind has limitations. The groups were not randomised. It is not clear whether the entire ability range was included in the study. The study seems to be as much concerned with reasoning skills as with argumentative writing.

Weight of evidence C (relevance of focus of study to review)
Medium: Again, it is not clear whether the conceptual focus is upon reasoning skills or upon written argument. It is not clear how or why skills acquired orally in a group situation are best measured through an individual written outcome - the topic of which, incidentally, could be construed as being a little biased in terms of gender and culture.

Weight of evidence D (overall weight of evidence)
Medium: This is a good study within its limits, which are acknowledged by the authors. However, the lack of randomisation, the weaknesses in the validity of the data-collection methods and the analysis - and of the overall design of the study, with a vocabulary test used as a baseline pre-test measure - all suggest that the study can only be of medium, or medium-to-low significance. There is the suggestion, too, that the authors wanted to prove that collaborative reasoning is a helpful approach in improving not only persuasive writing, but also reasoning skills more generally. Although they are cautious at the end, they are probably too optimistic about the transfer of argumentative (oral) skills via abstract schemas.

The authors themselves acknowledge the limits of their study. Although the data has been subjected to rigorous statistical analysis, the study's 'centre of gravity' is never clearly identified: the authors seem torn between exploring reasoning skills and measuring the use of argumentative strategies through the medium of writing. Their qualitative analysis of a sample of the students' essays seems very subjective. There is a lack of information to support the validity of the essay scoring procedures. More contextual information about the schools involved in the study is needed. The study is only 'quasi- experimental.' (p 173)

Troia GA, Graham S (2002) The effectiveness of a highly explicit, teacher-directed strategy instruction routine: changing the writing performance of students with learning disabilities. *Journal of Learning Disabilities* 35: 290-305

Country of study
USA

Age of learners
9-11: grades 4 and 5

Type of study
Researcher-manipulated evaluation: randomised controlled trial

Aims of study
To examine the effectiveness of a highly explicit, teacher directed instructional routine used to teach three planning strategies to 4th and 5th graders with learning disabilities

Summary of study design, including details of sample
This is an individually randomised controlled trial. The instruction was carried out in groups of two. There were 10 instructional pairs (5 per treatment condition). Participants were pre- and post-tested in persuasive writing.

N = 20 4th and 5th graders

Nature of intervention
Instructors modelled how to use the three strategies (goal setting, brainstorming, and organising) to perform several different types of tasks (including story writing), explaining how the strategies were adapted for each particular task and how they affected performance.

Instructors identified multiple tasks and situations for which students could use the strategies.

Students were given homework assignments in which they applied the strategies to activities other than storywriting. For these assignments, instructors provided students with advice on how to apply the strategies to these new tasks and gave them feedback on each completed assignment.

Data-collection instruments, including details of checks on reliability and validity
The essay probe sessions (pre-test and post-test) involved the examiner reading out two topics, which the student selected.

Paper and pencils were provided. The examiner supplied correct spellings when asked. All papers were collected at the end (scripts and plans). Any potential bias that mechanical factors such as handwriting or spelling might exert during the scoring process were 'eliminated' due to the typing out of all essays with errors corrected. The inter-rater reliability was 0.80 for both stories and essays. The average of the overall scores assigned by the two raters was used in all subsequent data analyses.

No feedback on the writing was given. Titles were read out. Spelling was supplied where necessary. Pairs of topics were selected from a larger pool and counterbalanced across students and testing conditions. The order of story and essay texts was similarly counterbalanced. (p 295)

Four procedures were used to establish that the treatments were carried out as intended: (1) the instructors were trained until they were 100% accurate in carrying out the procedures, (2) weekly staff meetings were held during the instruction period, (3) the first author randomly observed a third of the sessions - instructors not implementing 100% of the steps were debriefed afterwards, and (4) all sessions were recorded and a third were checked by an independent observer (97.7% of the required steps were found to have been implemented). (p. 296). At the start of each session, the examiner reviewed the key elements of the genre being assessed and the key strategies to use (p 294). The suitability of the topic 'prompts' (i.e. titles) was reported as established in an earlier study (De la Paz and Graham, 1997)(p 295).

Methods used to analyse data, including details of checks on reliability and validity
Significance tests for treatment effects (difference scores) for stories and essays, using T-tests for independent means.

Checks on reliability and validity not applicable.

Summary of results
For stories: Students who were taught to use the strategies wrote stories that were qualitatively better than those produced by their peers assigned to the process writing condition.

For essays: There were no significant differences between groups in post-test difference scores for essay quality or essay length. the post-test essays written by children in the strategy instruction group were slightly longer but of lower quality than their pre-test essay. In contrast, the post-test essays written by students in the process writing group improved slightly in overall quality, but were shorter in length compared with essays written prior to instruction.

Product: No significant differences between the groups for essay quality (p=0.18) or length (p=0.57). Post-test essays slightly longer and of lower quality.

Process: No significant group differences for essay planning time (p=0.79). 'Students in both groups (strategy and process) spent little or no time (less than one minute) on advance planning for their pre-test or post-test essays' (p 298).

Propositions were 0, as students wrote no initial plans.

Conclusions
The authors concluded that teaching students with LD three basic planning strategies via an explicit and highly teacher directed approach had a modest positive impact on story writing.

When children were asked to write a paper in an unstructured genre, persuasive essay writing, there were no differences either in the amount of time spent planning in advance or in the length and quality of the resulting essays.

The brainstorming had been designed specifically to make it less attached to a specific genre and more generalisable.

The fact that no 'strategy' children applied any of the three planning strategies to the persuasive essay contrasts with the findings in an earlier study (Troia et al., 1999), where the three LD students did apply the strategies they had been taught. (p 301)

'The generalizability of our findings to special and general education classrooms is unclear' (pp 301-302)

Weight of evidence A (trustworthiness in relation to study questions)
High

Weight of evidence B (appropriateness of research design and analysis)
High: It is high in the sense that it is an RCT which provides information about the degree to which elementary children with LD can transfer (or rather do transfer) instruction from one genre to another.

Weight of evidence C (relevance of focus of study to review)
Medium: Problem with persuasive writing measures.

Indirect instruction is relevant to part of the review. The type of writing probe used involves one of the types of writing involved in the review, although as the authors do not present more than one sample topic ('Should students be required to wear uniforms at school?' (p 295)), it is not possible to establish definitively how far the essays covered the range of genres covered by the review.

Weight of evidence D (overall weight of evidence)
High to medium: The only problematic aspect, apart from the sampling at school level, is the role of examiner coaching, but the available evidence (for planning) suggests that its impact was minimal.

Van Tassel-Baska J, Johnson D, Hughes C, Boyce L (1996) A study of language arts curriculum effectiveness with gifted learners. *Journal for the Education of the Gifted* 19: 461-480

Country of study
USA

Age of learners
9-12: grades 4, 5 and 6

Type of study
Researcher-manipulated evaluation: controlled trial

Aims of study
The purpose of this study was to test the cognitive impacts on elementary-level gifted students at 4th to 6th grades on a specifically-designed 40-hour LA curriculum unit organised around the Integrated Curriculum Model (ICM).

Summary of study design, including details of sample
Cluster trial (quasi experiment) with intervention and control; 3 outcome measures at pre- and post-test.

Unclear how the seven experimental (N = 100 pupils) and 3 control classes (N = 54 pupils) were selected for participation in the study and how they were allocated to exp and control conditions. Some of the classes were full classes; others were groups of 'pull out' students.

Nature of intervention
Curriculum package designed to:
1. Develop literary analysis and interpretation skills
2. Develop persuasive writing skills
3. Develop linguistic competency
Each of these goals was addressed through' explicit teaching of selected literature by using discussion approaches with pre-determined questions, writing models, and a self-study grammar packet'.

Data-collection instruments, including details of checks on reliability and validity
Three types of measure (p 465):
Literacy analysis and interpretation outcomes
Persuasive writing outcome
Linguistic competency outcome
Scoring rubrics given for two of the outcome measures (including the persuasive writing) in appendices.

Measures of reliability for reading assessment and persuasive writing were developed by researchers and piloted with relevant populations. Writing measure rubric

developed by Toulmin et al. (1984) and used in earlier studies of persuasive writing.

Measure of grammatical understanding developed by Thompson (1992) and piloted.

Use of a 'consensus model' of agreement among three raters to ensure interrater reliability.

No details of validity - as above, the measures were based on previous measures developed for previous studies and piloted.

Methods used to analyse data, including details of checks on reliability and validity
Analyses of covariance - differences on pre- and post-test measures for experimental and control groups; analysis of variance to ascertain post-test differences between groups on the persuasive writing task.

Effect sizes were computed for the experimental group only with tests of statistical significance.

Measures of reliability and validity not applicable - standard statistical tests.

Summary of results
The experimental groups improved significantly in all three dimensions of assessment.

Pre- to post-test effect size for persuasive writing was 0.99 and statistically significant; the control groups (pre- to post-test) did not show significant growth in any of the LA assessed by the study.

Conclusions
The data from the current study lends further credence to the capacity of students to develop persuasive essays. Students in the sample experienced the greatest difficulty with warrants, providing sufficient elaboration of their argument to be persuasive.

The study provided preliminary evidence that focused and high powered and integrated curriculum intervention in the LA of even a relatively short duration can bring about important changes in student performance.

Weight of evidence A (trustworthiness in relation to study questions)
Low

Weight of evidence B (appropriateness of research design and analysis)
Medium

Weight of evidence C (relevance of focus of study to review)
Medium

Weight of evidence D (overall weight of evidence)
Low

VanTassel-Baska J, Zuo L, Avery LD, Little CA (2002) A curriculum study of gifted-student learning in the language arts. *Gifted Child Quarterly* 46: 30-44

Country of study
USA

Age of learners
7-14: grades 2 to 8

Type of study
Researcher-manipulated evaluation: controlled trial

Aims of study
To assess gifted student learning outcomes as a result of using a specially designed LA curriculum that wedded

the standards work to differentiation features of a gifted curriculum.

Summary of study design, including details of sample
'Quasi experimental design' but no information given about allocation into experimental and comparison groups. Pre- and post-tests given in literary analysis and persuasive writing.

46 schools (N=2,189), but not clear how many in experimental classes and how many in comparison classes

Nature of intervention
Four of a series of six units having six goals one of which was to develop persuasive writing skills, based on theories of discourse and argument.

Data-collection instruments, including details of checks on reliability and validity
A performance-based persuasive writing test that asked the students to develop an argument to support whether or not they would require all the students in their grade to read the given selection

Rater inter-reliability for scoring the instrument mentioned but not specifically reported for this study, but rather 'has been reported' elsewhere.

A well understood scoring system was used (based on Toulmin, 1958) to address validity.

Methods used to analyse data, including details of checks on reliability and validity
ANCOVA was used for comparisons across groups to answer the question of whether the treatment and comparison groups were significantly different in their post-test performance after controlling for pre-test differences.

Item analysis.

No details of reliability.

Validity addressed by checking requirements of ANCOVA; altering procedure and alpha level; removing certain students from the analysis.

Summary of results
'There was a statistically significant difference between the experimental and comparison groups on the post-test after adjusting for pre-test differences between the groups using ANCOVA - for persuasive writing the effect size was large' (p 36).

Conclusions
'The data suggested that the use of the William and Mary language arts units produces significant and important gains for gifted learners in key aspects of the language arts as assessed by demonstration of high-level thinking on performance-based measures.'

Weight of evidence A (trustworthiness in relation to study questions)
Low

Weight of evidence B (appropriateness of research design and analysis)
Medium

Weight of evidence C (relevance of focus of study to review)
Medium

Weight of evidence D (overall weight of evidence)
Low to medium

Yeh S (1998) Empowering education: teaching argumentative writing to cultural minority middle-school students. *Research in the Teaching of English* 33: 49-83

Country of study
USA

Age of learners
12-13: Grade 7

Type of study
Researcher-manipulated evaluation: controlled trial

Aims of study
To investigate the effectiveness of two heuristics based on Toulmin's (1958) model of argument and classical rhetoric for helping cultural minority middle school students in two different schools to write argumentative essays (p 49).

Summary of study design, including details of sample
This was a clustered trial using quasi-random experimental and case study research methods. Two schools were chosen in the San Francisco Bay area. In each, two classes were assigned, one as the experimental group and the other as the comparison group. Pre- and post-tests were administered to all participants.

An investigation was made into the effectiveness of two heuristics-based models of argument using quasi experimental and case study methods. A non-equivalent (pre-test and post-test) comparison group design (two schools, four classes, two conditions: control and experimental) was combined with case studies of 36 randomly selected students stratified by treatment group, classroom and three reading achievement levels.

Total number of students = 116

Nature of intervention
The interventions were heuristics: that is, plans and scaffolds for writing argument or 'devices to teach students a pattern of thought'. These were based on Toulmin's (1958) model of argument, which defined claims or propositions on the one hand, and supporting evidence or grounds on the other, as the basic elements of an argument (this is called 'thesis-support' in the article).

The two particular heuristics were a 'pyramid' closely modelled on Toulmin's model, with the claim or opinion supported by evidence; and by warrants and backing that make the connection between the opinion and the evidence valid. The second heuristic was a 'bridge', based on classical rhetoric (and most immediately on Fulkerson's 1996 work) in which the reason for a position was connected to an opinion by facts, 'if/thens' and values.

Data-collection instruments, including details of checks on reliability and validity
(1) Pre-test and post-test data taken before and after instruction to testing reading comprehension and procedural knowledge for writing arguments. Post-test included assessing use of the bridge and pyramid heuristics.

(2) Transcripts of interview with 36 random selected case study participants, including copies of all writing/essays produced during the six-week instructional phase of the study

(3) Survey responses administered to all students about their language and discourse patterns at home to assess use of thesis-support style argumentation

(4) Videotapes, audiotapes and observational data collected on a daily basis in each teacher before, during and after the six-week period of instruction.

The fourth set of data was used 'to monitor implementation of the treatments, collect information on contextual variables that could account for differences in outcomes, and frame data analyses and interpretations' (p 61).

Validity is addressed principally by using Toulmin and Fulkerson as models behind the design of the interventions/heuristics.

Methods used to analyse data, including details of checks on reliability and validity

The methods used to analyse the data are extensively described (pp 63-65). Essentially, inter-rater reliability was used against a scoring rubric that was validated for the purpose. The three minority sub-groups were aggregated together. Two analyses 'assessed students' knowledge of criteria and strategies for argumentative writing' (p 64). Essays using the pyramid heuristic were analysed according to categories of argumentative coding (claim, reason, etc.).

In summary:

(1) Pre- and post-test essays were typed and graded by one ETS (educational testing service) graders on three dimensions - development, voice and conventions - according to the six-point scoring rubrics. Gain scores were calculated.

(2) The transcripts of the 36 interviews were scored using a 6-point rubric based on the development scale for assessing students' knowledge of criteria and strategies. The case study data were assigned to two categories: (a) exhibits knowledge of claim-support structure, and (b) knowledge of strategies. To assess the pyramid heuristics each experimental post-test essay was parsed and phrases were coded as main claim, reason, supporting fact and value statement.

(3) Questionnaire data asking how often each student have his/her opinion and reasons when talking at home was tallied, converted to days per week using midpoint of each range of values and averaged.

Numerical data included means, standard deviations, p and f scores (stepwise analysis of variance).

Reliability is addressed principally through the inter-rater reliability mechanism.

Validity is addressed through reference to categories derived from Toulmin, Fulkerson et al.

Summary of results

Gain scores were higher in the experimental groups than in the control groups as far as argumentative development and voice were concerned, but not significantly higher for conventions. Development gain scores for cultural minorities were significantly higher in the experimental groups, but only marginally higher for the White students. The same is true for the results on voice.

From the survey (questionnaire) results, it appeared that Hispanic- and African-Americans were less aware of the thesis-support model of argumentation than White students (Asian-Americans were excluded from this finding because of the small sample).

The results also suggest that 'the effect of ethnicity is not due to the degree of familiarity with thesis-support argumentation or, alternatively, that the self-report measure of familiarity was inadequate' (p 67).

However, the author suggests that a more balanced sample of White and minority ethnic students is required to confirm these findings.

Conclusions

The author concludes that the heuristics made a difference between the experimental and control groups and 'in sum, explicitly teaching heuristics for thinking about the process of constructing arguments led to significant, albeit modest, gains in development and voice of essays written by minority students in the experimental group' (p 77).

The findings thus suggest that combining explicit and immersion (process) approaches for teaching argumentative writing are important, especially for minority ethnic groups.

Weight of evidence A (trustworthiness in relation to study questions)

Medium to low: Within their limits, the results can be trusted. There is a reasonable degree of reliability and the test and analyses are reasonably valid. The study is meticulously reported on all counts. However, teacher effect and selection bias are not taken into account.

Weight of evidence B (appropriateness of research design and analysis)

Medium to low: A randomised controlled trial with a balanced sample of different ethnic groups would have improved the overall research design. The fact that only two teachers participated in the trial, each teaching one experimental and one control group respectively, means that the possibility of teacher effects and selection bias was strong. The design of the study is not of a very high standard.

Weight of evidence C (relevance of focus of study to review)

Medium to high: Although the sample was predominantly ethnic minorities, the outcomes were relevant.

Weight of evidence D (overall weight of evidence)

Medium: This is a reasonably reliable, well reported study that draws on a range of data and is relevant to the research question. However, it is weakened by the clustered rather than individual basis of the quasi-random grouping, and the imbalances in the actual sample. Also, the potential for teacher effects and selection bias to influence the outcomes of the trial need to be taken into account.

The results of this systematic review are available in four formats:

SUMMARY	Explains the purpose of the review and the main messages from the research evidence
REPORT	Describes the background and the findings of the review(s) but without full technical details of the methods used
TECHNICAL REPORT	Includes the background, main findings, and full technical details of the review
DATABASES	Access to codings describing each research study included in the review

These can be downloaded or accessed at **http://eppi.ioe.ac.uk/reel/**

The **Evidence for Policy and Practice Information and Co-ordinating Centre** (EPPI-Centre) is part of the Social Science Research Unit (SSRU), Institute of Education, University of London.

The EPPI-Centre was established in 1993 to address the need for a systematic approach to the organisation and review of evidence-based work on social interventions. The work and publications of the Centre engage health and education policy makers, practitioners and service users in discussions about how researchers can make their work more relevant and how to use research findings.

Founded in 1990, the Social Science Research Unit (SSRU) is based at the Institute of Education, University of London. Our mission is to engage in and otherwise promote rigorous, ethical and participative social research as well as to support evidence-informed public policy and practice across a range of domains including education, health and welfare, guided by a concern for human rights, social justice and the development of human potential.

The views expressed in this work are those of the authors and do not necessarily reflect the views of the funder. All errors and omissions remain those of the authors.

This document is available in a range of accessible formats including large print. Please contact the Institute of Education for assistance:

telephone: +44 (0)20 7947 9556   email: info@ioe.ac.uk