The effectiveness of information and communication technology on the learning of written English for 5- to 16-year-olds

Richard Andrews, Allison Freeman, Dan Hou, Nick McGuinn, Alison Robinson and Judy Zhu

The authors are members of the English Review Group, one of a number of systematic research review groups working with the Evidence for Policy and Practice Information and Co-ordinating (EPPI) Centre. The EPPI-Centre has received funding from the Department for Education and Skills as part of an attempt to increase the evidence base for education in England and Wales. The particular project reported here was funded by the Teacher Training Agency working in collaboration with the EPPI-Centre. Richard Andrews is co-coordinator of the English Review Group (with Carole Torgerson), Dan Hou and Judy Zhu are masters and research students respectively in the Department of Educational Studies at the University of York, where Nick McGuinn is a senior lecturer, Allison Freeman is a research assistant, and Alison Robinson is an information officer. Address for correspondence: Alison Robinson, Department of Educational Studies, University of York, York YO10 5DD. Telephone: +44 1904 433000; fax: +44 1904 433459; email: ar31@york.ac.uk

Abstract
The last few years have seen an increase in research studies on the impact and effectiveness of information and communication technologies (ICTs) in the teaching and learning of English as a school subject. It is against that research background and against recent developments in policy and practice in the UK that the present systematic review of the effectiveness of different ICTs in the teaching and learning of English has been undertaken. The aim of this review was to shed light on whether ICTs are effective in the teaching and learning of English for 5- to 16-year-olds. A total of 2103 papers were found in the initial search of studies published between 1998 and 2003 on the topic of the review. An in-depth review on the effectiveness of ICT in the teaching and learning of written composition in English concentrated on nine studies. As eight of the nine studies were judged to be of medium weight of evidence and were also different from each other in nature, it was not possible to arrive at a clear answer to our in-depth research question. Rather, we wish to report that the field is in a pre-paradigmatic state where definitions of English, literacy and ICT are still relatively unclear, and where the causal and/or reciprocal relationships between them have yet to be fully theorised.
Introduction

Following a series of systematic reviews on the impact of information and communication technology (ICT) on literacy learning (Andrews et al., 2002; Burn & Leach, 2004; Locke & Andrews, 2004; Low & Beverton, 2004; Torgerson & Zhu, 2003) and a book which brought them together with a critical commentary and a chapter on methodology (Andrews, 2004a), the English Review Team at the University of York wished to move beyond reviews of studies of the impact of ICT on literacy learning to a specific study of the effectiveness of ICT in teaching and learning situations with regard to written composition, thus, a more focussed study with a particular research question in mind. The research question for the review reported here was ‘What is the evidence for the effectiveness of different ICTs in the teaching and learning of English (written composition), 5–16?’ Our earlier studies on impact, based on a review of research published since 1990, had revealed a mixed picture. In particular, we had been interested in the impact of ICT on learners rather than on classrooms, teaching, institutions or curricula. We found out, among other things, that evidence for the impact of networked ICT on literacy learning was in short supply: that literary studies had been affected, particularly in the provision of software to mediate between the reader and the book, but that motivation via ICT was often short-lived (and in some cases, in the medium term, demotivating); that moving image ICT had continued to foreground the reading of the moving image at the expense of the composition of moving image texts. In the one review in that set of reviews between 2002 and 2004 that had focussed on effectiveness (Torgerson & Zhu, 2003), the results were inconclusive (see Conclusion for further discussion). Part of the problem in the many studies that we have examined was that researchers had either looked for the impact of ICT on pre-ICT literacy or had undertaken small-scale or poorly executed and/or reported research.

In order to look more closely at the causal links between ICT and literacy development, we wish to intensify the examination of the relationship between ICT and the learning/teaching of literacy to focus on a study of effectiveness. Effectiveness, as opposed to impact, assumes a strictly causal connection, and is best measured via randomised controlled trials (RCTs) and non-randomised controlled trials (CTs). For the purposes of the review that we report here, we took research literature published between 1998 and 2003 in an attempt both to limit the scope of the review and to address the most recent research in the field.

Method

The method section will be necessarily brief, as a full account is given in the technical report of the review (Andrews et al., 2004). The systematic review (both the map and the in-depth study) approach that was used is set out in guidelines and tools devised by the EPPI-Centre (EPPI-Centre 2002a, b, c). In short, a protocol or plan for the research was drafted, including a provisional research question for the initial map of research in the field. Exclusion and inclusion criteria for the literature search were written. The protocol was peer-reviewed, revised and then published on the Research Evidence in Education website (http://eppi.ioe.ac.uk/reel). Research papers were searched for, identified, screened for relevance and then keyworded to create an initial database. A map
of research studies in the field was generated. From the map, one area of research was identified for in-depth review: written composition. Papers in this area were data extracted and assessed for quality and weight of evidence with respect to the research question. A narrative synthesis of the results was produced.

It is worth highlighting here the initial stages of the review, as they indicate the breadth and scope of the study we have undertaken. Discussion here of the early stages will enable us to concentrate on the results on the in-depth review in the Results section that follows. A total of 2103 papers were found in the initial search of studies published between 1998 and 2003 on the topic of the review. Of these, 56 met the inclusion criteria for the review. Of the 56 papers, 14 were reviews of research and 42 were reported primary research. Because four of the primary research papers contained more than one study, a total of 14 reviews and 53 studies were examined.

Nine of the 14 reviews were systematic reviews, and the countries of origin of the 14 reviews were the UK (7) and the US (7). They also divided equally between a focus on reading, writing and other aspects of literacy. The general consensus of the reviews with regard to written composition was that computer-assisted instruction (CAI) or learning and word processing appeared to have a beneficial effect on students’ written composition, but only when combined with strategic instruction.

Of the 53 primary studies, 36 were from the US, 15 were from the UK, and one each was from Australia and Canada. The majority of the studies focussed on learners between the ages of 5 and 10 in a primary or elementary school setting; two thirds involved learners of both sexes. Again, these studies were evenly divided between an emphasis on reading, writing or other aspects of literacy. The principal focus in terms of ICT across the 53 primary studies was on CAI or computer-aided learning (CAL), and on software. Almost two thirds of the studies on writing focussed on composition. More specifically, in terms of those 20 studies that focussed on written composition, CAI/CAL software and multimedia were the most popular ICT interventions, each with six studies; five studies investigated word processing.

In the end, we came down to an in-depth review of nine studies in an attempt to answer our research question. A study had to focus on ICT and written composition to be included in the in-depth review. The reasons for focussing on this topic were partly pragmatic and partly related to policy, research and practice. Pragmatically, we needed to concentrate on an aspect of English that had not been heavily researched (unlike reading) and that might give us evidence of different outcomes according to the kind of ICT used. In terms of policy, research and practice, we felt that a focus on written composition would provide clear evidence of outcome and of the effectiveness of ICT interventions. In seeking good quality evidence for effectiveness, the inclusion criteria were further refined by the addition of a filter that limited included studies to RCTs and CTs with adequate sample size and quantified outcome measures. Sample size is relevant because, even if there is a control group, the study cannot be regarded as sufficiently valid or reliable to answer questions of effectiveness if the total sample is fewer
than 10. We restricted our attention to trials with quantified outcome measures as these are acknowledged as the best measures of effectiveness.

Our full criteria for the in-depth review, then, were as follows: the studies had to focus on ICT and written composition, be an RCT or CT, have adequate sample size to balance all the covariates and have quantified outcome measures. The rigour of these criteria is essential to systematic reviews, and the particular nature of these sets was determined in large measure by our aim to find evidence of effectiveness—hence, the concentration on experimental trials.

Results
First, we present the nine included studies (Table 1):

All nine studies selected for in-depth review were carried out or were assumed by implication to have been carried out in the US. In each study, the main pedagogical focus was learning; it was impossible to identify in any of the studies a split between teaching and learning, hence the title of this paper (despite our intention to focus on learning and teaching). As summarised in Table 2, the majority of studies \((n = 8)\) involved learners between the ages of 11 and 16. Three studies involved learners aged between 5 and 10 and two included both age groups. This differs significantly from the systematic map, in which the proportion of studies involving learners of primary school age was much higher.

Table 1: Studies included in in-depth review

<table>
<thead>
<tr>
<th>Author(s), year, title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cramer, S. &amp; Smith, A. (2002). Technology’s impact on student writing at the middle school level</td>
</tr>
<tr>
<td>Lowther, D. L., Ross, S. M. &amp; Morrison, G. M. (2003). When each one has one: The influences on teaching strategies and student achievement of using laptops in the classroom</td>
</tr>
<tr>
<td>Rowley, K., Carlson, P. &amp; Miller, T. (1998). A cognitive technology to teach composition skills: four studies with the R-WISE writing tutor (Study C)</td>
</tr>
</tbody>
</table>
Three types of ICT intervention were used in the studies selected for in-depth review: CAI/CAL, word processing and multimedia. As shown in Table 3, each study focussed exclusively on one type of ICT. The proportion of studies that used word processing is high compared with the proportion of studies in the map, but this is not surprising, given that the focus of the in-depth review is written composition.

All the studies in the in-depth review were examined in order to determine the nature of the synthesis to be conducted for the review. It was not felt to be appropriate to conduct a meta-analysis because the samples (ages, learner characteristics, etc.), interventions and outcome measures were too heterogeneous. Therefore, a narrative synthesis was undertaken.

Of the nine studies included in the in-depth review, eight were judged to be of medium weight of evidence in answering the research question for this particular review. All nine studies were CTs in one form or another, but only in one of the studies (Lewis, Ashton, Haapa, Kieley & Fielden, 1999) was there an element of randomisation. This section discusses these subsequently. The synthesis starts with a discussion of Lewis et al
As Lewis et al. (1999) state, ‘the primary goal of [their] study was to examine the impact of word processing tools combined with effective instruction on editing and revising performance’ in English (p. 87), specifically on the effectiveness of spelling and grammar checkers as tools for improving the quality and accuracy of the writing of students with learning disabilities. The study type is that of an RCT within a controlled trial, with A, B and C teacher groups randomly allocated. This is an unusual study type in that randomisation is used in a particular way: not to increase reliability between the control and experimental groups, but to minimise the potential influence of teachers upon the learners in the conduct of the study. As part of the pretest and posttest control group design, 106 students with learning disabilities made up three experimental groups and one control group. A further control group with no disabilities was established. Both control groups worked with only paper and pencil, as opposed to the experimental groups who worked with a Write This Way word-processing programme. The study was rated high to medium in terms of weight of evidence, in that there was a tight focus on quantifiable outcomes in terms of writing (such as the number of ‘mechanic’ errors); use of control and comparison groups; and an application of established data collection procedures. However, the number of teachers in each of the A, B and C groups was not stated.

Table 4: Individual weights of evidence of studies selected for in-depth review

<table>
<thead>
<tr>
<th>Paper</th>
<th>Weight of evidence A (trustworthiness in relation to study question)</th>
<th>Weight of evidence B (appropriateness of research design and analysis)</th>
<th>Weight of evidence C (relevance of focus of study to review)</th>
<th>Weight of evidence D (overall weight of evidence)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lewis, Ashton, Haapa, Kieley &amp; Fielden (1999)</td>
<td>High to medium</td>
<td>High to medium</td>
<td>High</td>
<td>High to medium</td>
</tr>
<tr>
<td>Barrera, Rule &amp; Diemart (2001)</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
</tr>
<tr>
<td>Cramer &amp; Smith (2002)</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
</tr>
<tr>
<td>Fan &amp; Orey (2001)</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
</tr>
<tr>
<td>Lowther, Ross &amp; Morrison (2003)</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
</tr>
</tbody>
</table>
The results of the Lewis et al (1999) study show little difference between the treatment/experimental groups of ‘learning disabled’ students in terms of writing quality. There were differences, however, in those groups that used spelling and grammar checkers, in that the number of errors decreased from pretest to posttest for all experimental groups. Significantly, the decrease happened at a faster rate for students with learning disabilities who were using the software. Furthermore, the software gave those students who used it more self-esteem in themselves as writers; 86% of students agreed that Write This Way was a helpful tool for writing, and 64% said they were better writers on the computer than with paper and a pencil.

The remaining studies, for various reasons described in detail in the full report, were afforded medium weight of evidence overall. This means that only Lewis et al (1999) can be depended on for a relatively authoritative answer to the research question we set ourselves (with the caveats previously stated), and that the following studies, while shedding light on the problem from various angles, are neither robust nor dependable enough to provide solid evidence.

Barrera, Rule and Diemart (2001) sought to build on recent investigations of computer use by pre-school children by investigating a mixed-ability class of first-grade students in a semirural elementary school. A relatively small sample of students (18) was given alternating interventions and activities over a 6-month period, and their performance was measured to assess the differential effects of computer word processing and handwriting of assignments. The measures included number of words and sentences written, and also on- and off-task behaviour during writing periods (to test concentration on the task in hand). Computer use resulted in significantly more writing: on average, there were more sentences written per assignment and more words per assignment when using a computer than when assignments were handwritten. Off-task behaviour differences were not significant.

In the study by Cramer and Smith (2002), the aim was to find out how student writing in the areas of organisation, voice and/or ideas improves with involvement in technology-rich instruction. In a quasi-experimental study, with pre- and posttest design, a sample of 139 students was used, with 88 in the experimental group and 51 in the control group. The results were that changes in achievement did occur, but in predicted directions, they did not. ‘By the end of the year, there was no statistically significant differences in student writing abilities (p > .05)...’. The median score was the same for both conditions, and no different from the pretest scores in general. The ‘only differences in scores on the post-test were for sixth graders in the areas of organization and voice (p > .5)’ (p. 8). There are other small effects reported, but nothing of note for this review. The authors conclude that the research ‘does not provide evidence to support a yes vote’ (p. 8) with respect to its research question.

The purpose of Fan and Oreys (2001) study ‘was to examine the constructivist use of multimedia technology to improve students’ writing performance’ (p. 2) with a controlled trial, pre- and posttest writing samples and a sample of 47 seventh-grade
advanced language arts students. The 6-week experimental project did not improve compositional skills in pencil and paper essays; rather, ‘its power lay in its motivational effects rather than its effects on learning writing’ (p. 11). The authors conclude that more systematic investigation of motivational effects would be needed to gauge the extent to which motivation has been affected—and, we would add, sustained.

The principal aim of Lowther, Ross and Morrison (2003) was ‘to provide further insight into the degree to which school laptop programmes can influence students’ educational experiences and learning’. Two primary research questions were asked: Is teaching and student behaviour different in laptop compared to control classrooms? And do students achieve differently in laptop classrooms? The study focussed on fifth-, sixth- and seventh-grade students, and a number of evaluation measurement strategies were used. Results were highly significant with regard to the effect of using laptops on writing, particularly on content, organisation and style; the fourth aspect of writing capability that was measured (conventions) showed a slightly lower significance. However, the ex post facto design of the study, despite the presence of a control and experimental groups over 21 classrooms (the exact sample is unclear), reduces the overall reliability of the results.

Rowley, Carlson and Miller (1998) undertook a linked series of four studies, the broad aim of which, as a group, was ‘to evaluate and adapt the writing software Reading and Writing in a Supportive Environment (R-WISE) in an attempt to improve prose composition as a cognitive act with the help of computers’ (p. 260). R-WISE is a word-processing package based on an understanding that composing in writing is a cognitive, socially informed act that might go through a number of stages. It therefore tries to reflect contemporary understanding of the writing process rather than being a drill-based tool. The four separate studies took place sequentially, developing the software as well as measuring outcomes in composition along the way. A large sample was used for each of the studies.

The first study revealed that R-WISE helped most ninth graders in the sample to improve their compositional skills, but helped students with lower pretest scores the most. Specifically, ‘holistic score trends... suggest that the use of R-WISE effectively supplemented and extended classroom instruction in writing. Introducing R-WISE into ninth-grade English classes produced significant gains over and above traditional instruction’ (p. 272).

The second study showed similar results, in that on all measures, the treatment/experimental group gained about twice as much progress on compositional skills as the control group (again with a large sample). This time, higher-end students seemed to gain more than lower-end students, in contrast with the findings of the first study. The interesting difference between the two studies was that the second one compared the use of R-WISE with straightforward word-processing, rather than with handwritten composition. The novelty factor of the computer was thus reduced.

© 2006 The Authors. Journal compilation © 2006 British Educational Communications and Technology Agency.
Whereas the focus of the first two studies was on comparing tools, the third study was more instructional in focus, in that a comparison of ‘guided’ with ‘open’ modes of composition was made. The sample was 1122 students and 21 teachers. Using the same 6-point holistic scoring rubric as in the first two studies, the authors revealed that ‘groups using R-WISE in guided mode during the first semester then open mode during the second semester showed significant gains over groups using only guided for both semesters’ (p. 280). Importantly, ‘the teacher’s instructional style appeared to play some role in the effectiveness of R-WISE, with the students of those teachers who prefer social, independent, and neutral instructional styles receiving less benefit from R-WISE than students of teachers who prefer a conceptual instructional style’ (pp. 280–281).

The authors conclude that as ‘the level of prior knowledge that a student has of a domain has been demonstrated to be a critical factor in the effectiveness of learner-control in computer-based instructional systems’, then the fact that students in the sample did best when moving from guided mode into open mode seems to confirm such findings.

The fourth and last of the Rowley et al (1998) studies was a replication of previous studies, using R-WISE version 3.0. In other words, it attempted to see if results of previous studies held up to sustained enquiry. The results revealed that they did, i.e., there was superior performance by ninth-grade students using R-WISE on all measures, but the size of the difference remained small—as in earlier studies—and the students' pretest scores remained more predictive of posttest performance than the effect of the experiment. The authors also conclude that ‘to some degree the application of cognitive science to the development of an adaptive learning environment for writing skills is viable’ (p. 291), but they note that several unanswered questions remain, including the context of student aptitudes and the support of teachers' preferred instructional styles.

In summary, this was a heterogeneous set of studies with no possibility of meta-analysis. In arriving at a narrative synthesis in answer to our research question, we have judged all the studies (with the exception of Lewis et al, 1999) to be of medium weight of evidence in answering the question. All the studies suggest, rather than conclusively prove, the beginnings of an answer to the research question we set ourselves. Part of the difficulty in attempting a synthesis is that the studies are so different. Barrera et al (2001) concentrate on length of composition and on attitudes to writing; Cramer and Smith (2002) concentrate on organisation, ‘voice’ and ideas; Lowther et al (2003) focus on content, organisation, style and convention; Rowley et al’s (1998) four studies focus on the impact of a developing word-processing programme on student composition. At least all of these studies, and that of Lewis et al take ICT as word processing on a computer, whereas Fan and Orey (2001) look at multimedia technology more broadly. In short, there is no consistent view across the set of studies on the nature of the written composing process, nor on ICTs themselves. We are thus dealing with two variables—ICT and written composition—which require further definition. While there has been at least four decades of research on writing processes and perhaps two on ICT, the field is still in a preparadigmatic state where researchers pay too little attention to defining
the (shifting) terms with which they are operating. In such a stage of development in research in a particular field, syntheses remain largely heterogeneous; we are thus unable to make confident comparisons between the effectiveness of different ICTs on learning in English for 5- to 16-year-olds.

**Conclusions**

The quality and variability of these results suggest that further research is needed to answer the specific question we set ourselves. There has yet to be a high-quality, large-scale RCT in the field that would answer the question about the effectiveness of ICT in assisting written composition for 5- to 16-year-olds. In this sense, the review has come to the same conclusion as Torgerson and Zhu (2003), a more extensive in-depth review on evidence for the effectiveness of ICT on literacy learning. As indicated in the earlier chapters of this review, Torgerson and Zhu systematically reviewed studies from 1990 to 2002 (rather than the more recent 5-year span of the present review); they were also able to conduct a meta-analysis of the results of two studies (which this review has been unable to undertake because of the nature of the studies examined). They had found that previous systematic reviews were equivocal on the effectiveness of ICT on literacy learning in English for 5- to 16-year-olds, and their own review confirmed the inconclusive nature of studies on this topic. They found that research had suggested that ICT was neither beneficial nor harmful with respect to its impact on literacy learning; they recommend that large-scale randomised trials be undertaken in the field to work towards a clearer answer to the question of effectiveness. As far as we know, no such study has yet been undertaken.

Perhaps one of the most exciting implications to emerge from the attempted synthesis of studies in the in-depth review is that the field of research in ICT and literacy/English is still in its infancy; it is in a preparadigmatic state. That is to say, we have had a number of studies that purport to try to answer the question of the effectiveness or impact of ICT on literacy development, and fewer studies that try to explore the reciprocal relationship between the two. What we are yet to see is a coherent theoretical account of the field that builds on the work of the New London Group in the mid-1990s (see Cope & Kalantzis, 2000) and that provides a clear and workable framework for further research, with particular emphasis on the use of ICT. Andrews (2004b) is currently working on such an account, with notions of the reciprocal coevolution of ICT and literacy.

Finally, it may be the case that no amount of prospective or ex post facto research is going to solve the problem of how best to bring research, policy and practice together to improve student learning in English. Rather than more and better studies (which is a laudable aim in itself), one of the implications of the recent studies on the impact and effectiveness of ICT on literacy learning in English is that we also need a new kind of research: one that is at the cutting edge of interface design in the service of learning. Such research will look more like research and development. We hope that in this study, we have contributed to a firmer foundation for such advances in human knowledge and pedagogical practice.
Acknowledgements
The authors would like to thank Carole Torgerson for her support throughout the project, in particular on questions of methodology; staff at the EPPI-Centre at the Institute of Education, University of London (particularly Diana Elbourne and Katy Sutcliffe) for their help during the review; and Marilyn Leask and Cherry White at the Teacher Training Agency for funding the review, and also for their intellectual interest in the development of the project.

References
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, Institute of Education.
EPPI-Centre (2002b). EPPI-Reviewer® (Version 0.9.7). London: EPPI-Centre, Social Science Research Unit, Institute of Education.
EPPI-Centre (2002c). Guidelines for extracting data and quality assessing primary studies in educational research (Version 0.9.7). London: EPPI-Centre, Social Science Research Unit, Institute of Education.

