NRDC

National Research and Development Centre
for adult literacy and numeracy

Research Review

Adult literacy and numeracy interventions and outcomes: a review of controlled trials

Carole Torgerson, Greg Brooks, Jill Porthouse,
Maxine Burton, Alison Robinson,
Kath Wright, Ian Watt

March 2004

Adult literacy and numeracy interventions and outcomes: a review of controlled trials

Carole Torgerson, Greg Brooks, Jill Porthouse, Maxine Burton, Alison Robinson, Kath Wright, Ian Watt

CONTENTS

Preface

The reviews contained in this report are linked reviews undertaken by the National Research and Development Centre for Adult Literacy and Numeracy: (1) expert review; (2) scoping review; (3) systematic review of randomised controlled trials (SRCTs) and (4) systematic review of controlled trials (CTs). These reviews have been designed to remedy the current lack of systematic reviews of all the randomised controlled trials (RCTs) and other controlled trials (CTs) in the area of adult literacy and numeracy. Such reviews meet two aims: firstly, to assess whether any rigorous evaluations have been undertaken and secondly, to inform the research agenda for future RCTs or CTs that may be used to evaluate interventions in adult literacy and numeracy.

Project team and authorship

Carole Torgerson, joint project director, University of York
Professor Greg Brooks, joint project director, University of Sheffield
Jill Porthouse, research fellow, University of York
Maxine Burton, research fellow, University of Manchester
Alison Robinson, research secretary, University of York
Kath Wright, independent information consultant
Professor Ian Watt, systematic reviews consultant, University of York
Jacquie Gillott, project secretary, University of Sheffield

Greg Brooks advised throughout on content, wrote the expert review and edited the final manuscript. Maxine Burton analysed the qualitative evidence from the RCTs and wrote section 3.9 and Appendix H of the systematic review of RCTs. Alison Robinson compiled and maintained the electronic databases for all reviews and provided administrative support. Kath Wright undertook the electronic searches. Ian Watt advised at all stages of the reviews and commented on drafts. The rest of this report is the work of Carole Torgerson and Jill Porthouse who undertook the screening, coding and analyses in the scoping and systematic reviews and jointly wrote the scoping review. Carole Torgerson wrote the protocols and the systematic reviews of the RCTs and the CTs.

Acknowledgments

We would like to thank the Adult Basic Skills Strategy Unit at the Department for Education and Skills, the government department that funds the National Research and Development Centre for Adult Literacy and Numeracy.

We also acknowledge the advice and support of the members of the expert panel who advised on the systematic review of randomised controlled trials – Dr Andrea Nelson and Dr Lewis Pike – and thank them for their valuable contributions at each stage of the review.

Summary

(1) Expert review

Background
This expert review was intended to complement the scoping review and the two systematic reviews, which constitute the rest of this report. Those reviews were designed to address very rigorously the question, 'What factors in teaching adult literacy and numeracy cause learners to make progress in learning?' and in order to do so considered only randomised and other controlled trials. As is frequently the outcome of systematic reviews in education, few clear pedagogical recommendations resulted. This expert review, therefore, considers a broader range of evidence.

Objective
The objective of this expert review is to provide a synopsis of research findings in the field of effective practice. The findings are categorised in this summary into

(1) those based on randomised and other controlled trials; (2) those based on correlational evidence; and (3) those based on researchers' and other experts' judgements.

Method
The review of research by Brooks et al. ['Assembling the Fragments: a Review of Research on Adult Basic Skills'] was taken as a baseline. Several more recent reviews, including and especially the systematic reviews in this report, were used to update the information. Those used, other than the systematic reviews here, are listed in the references to the full expert review.

Findings
(1) Findings based on randomised and other controlled trials:

- Receiving adult literacy and numeracy tuition produces more progress than not receiving it.
- Instruction using information and communications technology (ICT) cannot yet be shown to produce more learning than conventional instruction.
- Reciprocal teaching had positive effects on reading comprehension. This was the strongest single finding on pedagogy.
- A 'diagnostic prescriptive' approach had positive effects on reading comprehension, but not on word identification.
- For inmates at a United States prison, a 'community-building group process' accompanied by the SRA reading programme had a positive effect on reading.
- A 'modified numeracy' approach had positive effects on arithmetic.
- One study produced a negative finding. It investigated the effectiveness of auditory perception training on the reading ability of adult poor readers at a community college. The control group made more progress on a word identification test and it seems reasonable to conclude that auditory perception training was not effective.

(2) Findings based on correlational evidence:

■ Phonemic awareness and/or word analysis instruction may lead to increased achievement in other aspects of reading for adult beginning readers.

■ Fluency (meaning greater speed in reading aloud) may be taught to adult learners, and fluency practice may lead to increases in reading achievement.

■ Providing explicit instruction in reading comprehension strategies may lead to increased reading comprehension achievement.

■ Combining comprehension instruction with instruction in various other components of reading may lead to increased reading comprehension achievement.

■ Factors associated with better progress in reading made by adult learners in a large-scale study in England and Wales were: the tutors in an area having qualified teacher status; tutors having assistance in the classroom; and regular attendance by students.

■ In further education (FE), the provision of basic skills support reduces drop-out rates and increases completion rates.

■ The amount of instructional time learners need to receive to make educationally significant progress is estimated in the USA to be at least 100 hours of instruction to make progress equivalent to one grade level.

■ Two features of effective programmes in England and Wales are that they have high expectations of learners' achievements and that they enable learners to gain credit and accreditation for their learning and allow learners to move on to further study should they want to.

(3) Findings based on researchers' and other experts' judgements:

■ Effective programmes in England and Wales provide for the acquisition of skills in a range of contexts (meeting the motivation and interests of learners); deliver clearly structured teaching; produce a learning plan for each learner; regularly assess and review progress and adjust individual learning plans (ILPs) accordingly; and adjust the length of programme to the level of skills required.

■ The progress made by parents in family literacy and numeracy programmes appears to relate to several factors, including: the parents' commitment to improving their children's chances, clear goals, quality of staff and teaching, time-limited courses to focus achievement and joint (as well as separate) teaching sessions for parents and children.

■ Research suggests that students can build successfully on their informal mathematical knowledge to construct meaning from formal representations, although a clear relationship must exist between the two for this to happen.

■ Small-scale teacher-researcher studies in the USA on adults' multiple intelligences with regard to adult numeracy suggest that this approach may be worth pursuing.

■ Factors thought to be related to progress in workplace basic skills are all structural; no pedagogical factors appear in the literature.

■ A review of assessment instruments for adult literacy and numeracy found none suitable for use in NRDC's research programme.

(2) Scoping review

Background
This scoping review was designed to remedy the current lack of a review of all the randomised controlled trials (RCTs) and controlled trials (CTs) in the area of adult literacy and

numeracy. Such a review meets two aims: first, to assess whether any rigorous evaluations have been undertaken of interventions that could be implemented and secondly, to inform the research agenda for future RCTs or CTs that may be used to evaluate interventions in adult literacy and numeracy.

Objectives

The primary objectives of this scoping review were:

- to search for, locate and synthesise all the RCTs and CTs intended to evaluate interventions in adult literacy and/or numeracy and published between 1980 and 2002;
- from those RCTs and CTs to establish what conclusions, if any, could be drawn about the effectiveness of teaching; and
- to investigate whether or not there was any difference in the pooled effect size of all the RCTs compared with the pooled effect size of all the CTs where effect sizes could be calculated.

Methods

Systematic review methods were used throughout, in order to limit bias. Studies were systematically searched for and located. They were then screened for inclusion in the scoping review using pre-established inclusion and exclusion criteria. At this screening stage of the review two researchers (Carole Torgerson and Jill Porthouse) worked independently; they then conferred and agreed on all decisions. The included RCTs and CTs were data extracted and synthesised in mapping tables.

Results

A total of 4,555 potentially relevant papers were identified using the electronic and hand searches as described in Appendix D.

The 4,555 papers identified were double screened, using the inclusion/exclusion criteria established in the protocol. Of these, 4,387 papers were excluded at this first stage and 168 papers were sent for. Within this subset, 19 papers were unavailable or were not received. All papers received were re-screened on the basis of a reading of the full paper. At this (second) stage, 90 papers were excluded, leaving a total of 59 papers to be included in the review. The titles and abstracts of all the papers in the database are available as an EndNote library on request (ar31@york.ac.uk), together with the reasons for their inclusion or exclusion. The 59 included papers were coded and the database was fully annotated with the codes (by Alison Robinson). Ten reviews were identified (and included) through the electronic searches and through a contact. One RCT was identified from a hand search of the bibliographies of these included reviews.

Of the 59 studies that were included in the scoping review, ten were reviews or systematic reviews including one or more RCTs and/or CTs. Altogether there were 12 RCTs and 34 CTs. Of these 46 trials, nine RCTs and 27 CTs contained data; three RCTs and seven CTs contained no data. The trials containing no data were excluded from the scoping review; and the remaining 36 trials (nine RCTs and 27 CTs) were data extracted. The nine RCTs were independently double data extracted, quality appraised and summarised for the systematic review of RCTs. Nine of the CTs included sufficient data in order for the effect sizes to be calculated. The effect sizes in one CT (Dietrich, 1994) could be calculated after the author replied to a request to provide numbers in the intervention and control groups in the results section. These nine CTs were independently double data extracted, quality appraised and summarised for the systematic review of CTs.

Of the 36 adult basic education studies, 24 had only a literacy focus, two had only a numeracy focus, and ten examined both.

Of the 34 with a literacy focus, seven were RCTs and 27 were CTs. In most of these studies the intervention consisted of computer-assisted instruction (CAI – 18 studies). Five studies reported a positive effect for the intervention compared with the control group (the 'diagnostic prescriptive approach' to reading and the 'reciprocal teaching programme', both of which were RCTs and three CTs), and one study reported a positive effect for the control group, i.e. a negative effect for the intervention. Ten studies showed no difference (five RCTs and five CTs). The results of the remaining 18 studies were unclear because it was not possible for the reviewers to calculate an effect size. These results, however, should be interpreted with caution, as the quality of the CTs with insufficient data was not assessed.

Of the 34 studies, 33 were undertaken in the USA and one in the UK.

Of the 12 studies with a numeracy focus, four were RCTs and eight were CTs. In most of these studies the intervention consisted of CAI. One study reported a positive effect for the intervention compared with the control group (modified comprehension learning, which was an RCT), while six showed no difference (three RCTs and three CTs, of which all but one RCT were CAI interventions). The results of the remaining five studies were unclear (four CTs of CAI and one CT of a family literacy programme) because it was not possible for the reviewers to calculate an effect size. These results, however, should be interpreted with caution, as the quality of the CTs with insufficient data was not assessed.

Of the 12 studies, ten were undertaken in the USA, one in New Zealand and one in the UK. In the RCT that took place in the UK (Nicol and Anderson 2000) the effect of CAI on the numeracy skills of adults with 'mild learning disabilities' was assessed. The authors reported that there was no difference between the CAI group and the traditional teaching group.

Discussion
Out of a total of 46 trials (12 RCTs and 34 CTs) included in this scoping review only 36 contained data (nine RCTs and 27 CTs) and of those 36 trials only 18 (nine RCTs and nine CTs) either reported effect sizes or contained sufficient data for the reviewers to calculate effect sizes. Only six studies out of the total of 36 trials showed a statistically significant positive outcome for the intervention. Three of these studies were RCTs: Cheek and Lindsey, 1994 (literacy), Rich and Shepherd, 1993 (literacy) and Schrum, 1985 (numeracy). The other three positive studies were CTs: Gretes and Green, 1994 (literacy); Maclay and Askov, 1988 (literacy) and Roberts et al.1994 (literacy).

In addition to the positive studies, there was one with a negative finding, that is a significant effect in favour of the control group: Dietrich, 1994 (literacy).

All seven of these studies are examined in more detail later in this report.

Conclusions
This overview of all the experimental research in the fields of adult literacy and numeracy since 1980 has revealed only six studies out of a total of 36 trials showing a statistically significant positive outcome for the intervention.

(3) Systematic review of randomised controlled trials (RCTs)

Background

In February 2002, as part of its ***Skills for Life*** strategy (GB. DfEE, 2001), the British government established the National Research and Development Centre for Adult Literacy and Numeracy to contribute to the improvement of adult basic skills teaching and learning. One of the first questions asked within the Centre's research programme was 'What conclusions can be drawn about the effective teaching of adult literacy and numeracy from the most rigorously designed experiments?' 'The most rigorously designed experiments' were defined as randomised controlled trials (RCTs), but no systematic review of RCTs in adult literacy and numeracy had been carried out. This project was therefore designed to carry out such a review.

Objectives

The primary objectives of this review were:

- to locate, synthesise and quality appraise all the RCTs intended to evaluate interventions in adult literacy and/or numeracy and published between 1980 and 2002, and
- to establish what conclusions, if any, could be drawn from those RCTs about the effectiveness of teaching.

Other objectives were:

- to analyse what the same studies had to say about the wider benefits to the adult learner of participating in adult literacy and numeracy interventions;
- to draw methodological conclusions about the quality of the research covered; and
- to draw up a proposal for a well-designed RCT on a key question in the field.

Methods

Systematic review methods were made explicit in the review protocol and were used throughout the review, in order to limit bias. Studies were systematically searched for and located. They were then screened for inclusion in the mapping section of the review using pre-established inclusion and exclusion criteria. Finally, the included RCTs were data extracted, quality appraised and synthesised. At the screening and data extraction stages of the review two researchers (Carole Torgerson and Jill Porthouse) worked independently; they then conferred and agreed on all decisions. A third researcher (Maxine Burton) carried out a qualitative analysis of the RCTs included for information on the wider benefits of learning.

Results

Electronic and hand searches yielded 4,555 possibly relevant papers, but the great majority (4,387) were deemed irrelevant on scrutiny of the abstracts, so that full copies of 168 were sent for. Of these, 19 were not available, and 149 were received and read in full. A further 90 were excluded at this stage, so that 59 papers were included in the descriptive map of the research in the field. Of these, 34 were CTs, ten were reviews and three were RCTs with no results. 12 papers were included that contained nine RCTs. These were data extracted and quality appraised. Most of the studies had substantial methodological weaknesses. There was also evidence of publication bias, with studies showing a null or negative effect missing from the data set.

However, three RCTs included comparisons of teaching against a no teaching control group. Pooling these three in a meta-analysis showed a strong positive and statistically significant, effect on outcome. Two other studies examined the use of CAI on literacy among imprisoned adults. Pooling these two showed a modest but statistically non-significant benefit. The analysis of wider benefits yielded very little of value.

Discussion

There is a dearth of rigorous RCTs in the field of adult literacy and numeracy. Few of the identified studies were UK based and only one of the nine RCTs. The evidence was just sufficient to show that adult literacy and numeracy instruction does benefit learners. While intuitively obvious, this result had not been rigorously demonstrated before and is therefore significant. However, the studies were so heterogeneous that no conclusions could be drawn about the precise effect of particular instructional practices and no useful conclusions could be drawn from the evidence on wider benefits of learning.

Conclusions

There is some evidence that adult literacy and numeracy programmes are effective although this finding may be undermined by the presence of substantial publication bias. We recommend that a series of large, well-designed and well-conducted RCTs be undertaken.

(4) Systematic review of controlled trials (CTs)

Objectives

The primary objectives of this review were:

■ to locate, synthesise and quality appraise all the CTs intended to evaluate interventions in adult literacy and/or numeracy and published between 1980 and 2002 and
■ to establish what conclusions, if any, could be drawn from those CTs about the effectiveness of evaluated pedagogies.

Methods

Systematic review methods were used throughout the review, in order to limit bias. Studies were systematically searched for and located. They were then screened for inclusion in the mapping section of the review using pre-established inclusion and exclusion criteria. Finally, the included CTs were data extracted, quality appraised and synthesised. At the screening and data extraction stages of the review two researchers (Carole Torgerson and Jill Porthouse) worked independently; they then conferred and agreed on all decisions. No attempt was made to analyse the wider benefits of the CTs because inspection of the papers revealed even less of value than in the RCTs.

Results

Electronic and hand searches yielded 4,555 possibly relevant papers, but the great majority (4,387) were deemed irrelevant on scrutiny of the abstracts, so that full copies of 168 were sent for. Of these, 19 were not available and 149 were received and read in full. A further 90 were excluded at this stage, so that 59 papers were included in the descriptive map of the research in the field. Of these, 12 were papers that contained nine RCTs, three were randomised controlled trials with no results, ten were reviews and 34 were controlled trials. Of these 34 controlled trials, seven contained no data, 18 contained insufficient data for the effect sizes to be calculated and nine contained sufficient data for the reviewers to calculate

the effect sizes. These nine CTs were independently double data extracted and quality appraised. The quality of the trials was variable but many of the studies had methodological weaknesses. There was little evidence of publication bias.

Discussion

In this systematic review we found only nine relevant CTs of interventions for adult literacy or numeracy where reviewers were able to calculate the effect sizes from the data included in the studies. It is of concern that the few trials that have been undertaken tend to be of low methodological quality, which renders their conclusions unreliable.

As with the randomised trials the literature is dominated by one country: the USA. Only one study, from New Zealand, was conducted outside the USA. Therefore, even if the trials had been of high rigour it would have been difficult to extrapolate their findings outside North America given the lack of replication in other English-speaking nations.

Within the CT literature there is substantial heterogeneity among the included studies in terms both of the settings in which the studies took place and of the interventions being evaluated. This heterogeneity makes it difficult to draw either quantitative or qualitative conclusions about which particular forms of intervention are effective.

Conclusions

The controlled trials tended to be of poorer reporting quality than the RCTs. There have been few attempts to expose common adult literacy or numeracy programmes to rigorous evaluation and therefore, in terms of policy and practice, it is difficult to make any recommendations as to the type of adult education that should be supported. In contrast, however, the review does provide a strong steer for educational researchers. Because of the present inadequate evidence base, rigorously designed RCTs are required as a matter of urgency.

'Expert' review

Forms of evidence

What counts as evidence depends on what the research question is. If the research question is 'What factors in teaching cause learners to make progress in learning?', then it could be argued that controlled (especially randomised controlled) trials have the potential to provide the most robust evidence. This is because only these experimental methods purport to control all possible extraneous causes of measured progress.

The scoping review and the two systematic reviews provided later in this report apply this logic rigorously. They adopt the most up-to-date methods of searching for and then analysing randomised and other controlled trials, in this case in the fields of adult literacy and numeracy. In terms of reliable findings for pedagogy, these reviews may seem to have somewhat meagre outcomes – these are summarised in the next section. But in terms of clearing away less rigorous research they are a model of their kind and provide a useful service.

It would still be legitimate, however, to ask what findings can be derived from other forms of evidence addressing different research questions in the field, for example:

■ 'What factors in teaching adult literacy and numeracy are known to correlate with better progress in learning?'
■ 'What factors in teaching adult literacy and numeracy are thought by experts to correlate with better progress in learning?'
■ 'What evidence is there that ICT enables adults to make better progress?'
■ 'What works for adult basic skills in the workplace?'
■ 'How much instructional time do learners need to make educationally significant progress?'
■ 'What assessment instruments are there in the field and how good are they?'

This review attempts to summarise what is currently known on these issues.

Findings from randomised and other controlled trials

This section is based entirely on the systematic reviews in this report.

The systematic review of RCTs found just enough evidence (all of it from the USA) to demonstrate rigorously in a meta-analysis that receiving adult literacy and numeracy tuition does produce more progress than not receiving it. Though this finding is intuitively obvious, this is the first time it has been rigorously demonstrated. If the finding had not been positive, questions would have had to be raised about whether provision should continue.

A further meta-analysis comparing conventional instruction with instruction using ICT is dealt with in the relevant section below.

Among the studies covered in the systematic reviews were several that had individual positive findings. Taking single studies as reliable entails making the assumption that other studies on the same topic with null or negative results do not exist – a very hazardous assumption given the possibility of publication bias (on which see the relevant sections of both systematic reviews) and the presence of one definite negative finding among the controlled trials (see below). However, if this assumption is made and the single studies are taken as reliable at least until further studies show otherwise, then we can say that these individual trials provide three possible further findings for adult literacy and one for numeracy:

■ Reciprocal teaching had positive effects on reading comprehension (Rich and Shepherd, 1993). Of all the pedagogical findings in this report this is the one in which we have most confidence. According to our quality appraisal in the systematic review of RCTs, this was the highest quality study of all – it used 'intention to teach analysis' and 'blind' random allocation to intervention and control groups using random number tables.

Reciprocal teaching is a teaching approach first described by Palincsar (1982) and then further developed by Palincsar and Brown (Palincsar and Brown, 1984; Palincsar, 1986). They describe it as:

> *A procedure ... where teacher and student took turns leading a dialogue concerning sections of a text. Initially the teacher modelled the key activities of summarising (self-review), questioning (making up a question on the main idea), clarifying and predicting.*

The teacher thereby modelled activities: the students were encouraged to participate at whatever level they could. The teacher could then provide guidance and feedback at the appropriate level for each student. (Palincsar and Brown, 1984, p.124)

The four activities are seen as having two functions, 'comprehension-fostering and comprehension-monitoring' (Palincsar and Brown, 1984, p.121). Pupils are gradually encouraged to take over the teacher role as they gain confidence and the whole approach is predicated on the idea that poorer comprehenders can improve by being shown and explicitly understanding and adopting good comprehenders' strategies. There has been a large amount of research on the technique in North America, but scarcely any in the UK and very little at adult level. However, Rosenshine and Meister (1994) did a meta-analysis on the 16 most rigorous studies at school level and found an effect size of 0.32 on standardised tests.

- A 'diagnostic prescriptive' approach had positive effects on reading comprehension, but not on word identification (Cheek and Lindsey, 1994). The approach involved formal and informal diagnostic procedures to identify adults' strengths and weaknesses and using the diagnoses to develop individual educational prescriptions.

- For inmates at a US prison, a 'community-building group process' accompanied by the SRA reading programme had a positive effect on reading (Roberts et al. 1994). This appears to parallel the finding at school level in Britain that working on children's self-esteem and reading in parallel has definite potential (see Lawrence, D. (1988) and summary in Brooks, 2002).

- A 'modified numeracy' approach had positive effects on arithmetic (Schrum, 1985). Phase 1 of the approach involved relaxation training and other psychological techniques; Phases 2 and 3 consisted of self-directed mastery learning, that is, the learners decided what objectives to pursue and pursued until satisfied with their own achievement.

However, if these positive findings from single trials are accepted, then there is also one negative finding among them that should also be accepted. In this study there was a statistically significant advantage for the control group, that is, the intervention had the opposite of the intended effect:

- The aim of the study (Dietrich, 1994) was to investigate the effectiveness of auditory perception training on the reading ability of adult poor readers at a community college. The experimental group received a phonological skills approach, while the control group received a traditional metacognitive approach. The results showed no difference on an auditory perception test or a word naming test, but the control group made more progress on a word identification test. Since the auditory test results suggested that auditory perception training was ineffective, perhaps the negative result for word naming meant that the control group were making better use of the time available.

Factors that are known to correlate with better progress

A recent and highly detailed US review of research on reading in adult literacy (Kruidenier, 2002) lists dozens of findings. It is based on both adult- and school-level research and distinguishes between 'principles' based on 'more than one' experimental study and 'trends' based on only one study or on indirect evidence (mainly from school-level research). Here only the principles based on adult-level research are listed. They are reproduced verbatim, hence the tentative tone:

- Phonemic awareness and/or word analysis instruction may lead to increased achievement in other aspects of reading for adult beginning readers.
- Word analysis may be taught using approaches that include direct instruction in word analysis along with instruction in other aspects of reading.
- Fluency (in this context, 'fluency' means greater speed in reading aloud) may be taught to adult basic education students and fluency practice may lead to increases in reading achievement.
- Fluency may be taught using approaches that include the repeated reading of passages of text, words from texts and other text units.
- Providing explicit instruction in reading comprehension strategies may lead to increased reading comprehension achievement.
- Combining comprehension instruction with instruction in various other components of reading may lead to increased reading comprehension achievement.

Brooks et al. (2001b) carried out a large-scale study for the Basic Skills Agency (BSA) of the progress in literacy made by adult learners in England and Wales. Average progress in reading was slow and modest and in writing almost non-existent. Factors associated with better progress in reading (none were found for writing) were:

- all the tutors in an area having qualified teacher status;
- tutors having assistance in the classroom;
- regular attendance by students.

An earlier piece of research commissioned by the BSA (1997) found that, in further education, the provision of basic skills support reduced drop-out rates and increased completion rates.

The BSA summed up its views on programmes that are most effective in enabling adults to improve their basic skills in its booklet **Effective Basic Skills Teaching for Adults** (BSA, 2000, especially p.11). Of the seven factors listed, two seem to be strongly based on quantitative experimental evidence, namely that effective programmes:

- have high expectations of learners' achievements;
- enable learners to gain credit and accreditation for their learning and enable them to move into further study if they so desire.

Factors that are thought to correlate with better progress

The other five factors listed in the BSA booklet seem to be based less on quantitative research than on its own and other experts' views. These other factors were that effective programmes:

- provide for acquisition of skills in a range of contexts meeting the motivation and interests of learners;
- deliver clearly structured teaching;
- produce a learning plan for each learner;
- regularly assess and review progress and adjust individual learning plans (ILPs) accordingly;
- adjust the length of programme to the level of skills required.

Family literacy and numeracy do appear to enable parents to make good progress in their own

basic skills (Brooks et al. 1996, 1997, 1999; BSA, 1998). The researchers judged that the following factors contributed to the parents' progress:

■ the fact that participation was voluntary;
■ parents' commitment to improving their children's chances;
■ clear information about goals, including subsequent progression;
■ clear focus on literacy/numeracy development, regarding other benefits as 'bonuses';
■ nationally recognised accreditation of learning;
■ careful selection of courses and staff and quality of teaching;
■ time-limited courses to focus achievement;
■ support of key personnel within institutions (e.g. primary headteacher);
■ inclusion of talking and writing, as well as reading, in literacy schemes;
■ teaching sessions for parents only and children only, as well as together.

For adult numeracy another key text is Coben's (2003) NRDC literature review. Her findings on pedagogy (pp.116–7) are as follows:

■ Evidence on the impact of adult numeracy tuition is sparse and unreliable. Detailed critical studies of adult numeracy teaching and learning, including intervention studies, are required before it will be possible to delineate good practice in the light of evidence rather than aspiration.
■ Studies investigating teaching and learning of particular elements of mathematics suggest that some (e.g. fractions, decimals, ratio and proportion) are likely to prove more difficult than others for some students.
■ Errors and misconceptions learned in childhood are likely to persist into adulthood.
■ Assessment methods developed within the Realistic Mathematics approach in The Netherlands may be useful in England. Adult learners write and publish their own mathematical problems, using a process of generating ideas, drafting, peer and teacher review and redrafting. This approach has been found to facilitate learners in developing conceptual understandings of mathematical topics as well as their communication skills.
■ There is some evidence that aspects of language acquisition will develop when supplemented with conceptual tasks and activities that focus on the written and oral use of mathematical understandings.
■ Research suggests that students can build successfully on their informal knowledge to construct meaning from formal representations, although a clear relation must exist between the two for this to happen.
■ Small-scale teacher-researcher studies in the USA on adults' multiple intelligences with respect to adult numeracy suggest that this approach may be worth pursuing.

Little of what is summarised in this section says anything directly about pedagogy.

What evidence is there that ICT enables adults to make better progress?

Information and communications technology did show some benefit for learners at entry levels 2 and 3 in a study conducted for Ufi (formerly known as the University for Industry) (Mellar et al. 2001) – but this was a single-group pre-/post-test study. In our systematic review of RCTs (see below), a meta-analysis of the two relevant studies showed no benefit of ICT over conventional instruction. Apparently on the basis of the same two USA RCTs (both conducted, as it happens, in prisons), Kruidenier (2002) expressed the finding more

optimistically as 'In general, computer-assisted instruction (CAI) is at least as effective as non-CAI in increasing reading comprehension achievement'. The further analysis of non-randomised controlled trials again found no convincing evidence of benefits from ICT over conventional instruction.

In total, the two systematic reviews found three RCTs and 16 other controlled trials relevant to ICT and adult literacy and/or numeracy. All three RCTs (one of which was not used in the meta-analysis because its subjects were not prisoners) had non-significant results. Of the 16 other controlled trials, three had no clear result, seven were non-significant and four had at least one statistically significant finding in favour of ICT, but two had statistically significant results in favour of the traditionally taught control group, that is, against the use of ICT.

A previous systematic review comparing computer-assisted and traditional approaches in adult reading achievement included 21 studies dated between 1979 and 1995, and had a finding broadly similar to ours:

> 'Despite considerable variability in design sophistication and clarity of reporting, ten of the 21 studies found no statistically significant differences between the two methods, five failed to indicate significance, two showed mixed results on differing assessment measures, three showed significance favouring CAI and one showed significance favouring traditional methods.' (Rachal, 1995, p.239)

Although this was a systematic review, the dates for including studies differed from the dates for including studies in our systematic review, as did other inclusion and exclusion criteria (age, outcome measures, unpublished theses – we only included RCTs, etc.). However, 11 of the 21 studies are included in our review, which partially explains the similarity in the findings.

What works for adult basic skills in the workplace?

Brooks et al. (2001a) found no studies at all which had gathered performance (test) evidence on the impact of workplace basic skills provision and this still seems to be the case. Rather, all the evidence in the field appears to be based on participants' and researchers' judgements. Given this, the factors thought to be related to progress in workplace basic skills can be summarised as:

- consultation between and commitment of all stakeholders
- subsidised costs
- time allowance
- suitable training ethos
- 'company awareness'.

But it should be noted that none of these is a matter of pedagogy. What is probably the first rigorous study in this field in Britain was started at the Institute of Education, University of London in 2003.

How much instructional time do learners need to make educationally significant progress?

Comings (2003) summarised the US evidence on this. Learners need to attend at least 100 hours of instruction to make progress equivalent to one grade level (after 150 hours the probability of making this much progress is 75 per cent), but the average learner stays in provision for fewer than 70 hours in a year. As with Learning and Skills Council retention figures in England, those who enrol but leave after just a few hours' attendance are not counted. For England and Wales, the Basic Skills Agency provided estimates of the variable amounts of tuition time learners take to move up one level of the (old) BSA Standards, with learners at lower levels taking longer on average (see the Table in Effective Basic Skills Teaching for Adults, BSA, 2000, p.19). No evidence is cited, but the figures seem to be based on the US research plus expert opinion.

Assessment instruments

In addition to the systematic reviews in this report and the review of numeracy research, NRDC has also carried out a review of assessment instruments for adult literacy and numeracy (Brooks, Heath and Pollard, 2004). A total of 15 instruments were analysed, that is virtually all the formal instruments used in Britain in the last 10 years, for their suitability for use in NRDC's research, against a range of criteria. The three most crucial criteria were that the instruments needed to be secure (unpublished), aligned with the National Standards and Curricula but also capable of providing finer differentiation and have parallel forms. None of the existing instruments met all three of these criteria. NRDC has therefore commissioned the National Foundation for Educational Research to develop a new instrument for adult literacy and there are plans to develop an instrument for numeracy.

Conclusions

A fair conclusion would be that, even though this analysis goes beyond controlled trials, it still illustrates the thinness of the research base. Yet it is already less exiguous than when Brooks et al. (2001a) carried out their review. The soundest findings can be summarised as follows:

■ Participation in adult literacy and numeracy provision does enable learners to make some progress.
■ Some hints on effective pedagogy can be derived from individual controlled trials.
■ Others (for reading) are suggested by the US research review.
■ A few other factors associated with greater progress in reading, for example regular attendance by students, seem well established by quantitative research.
■ Scarcely any factors associated with greater progress in numeracy are securely established and none for writing.
■ Though using ICT is undoubtedly an essential skill, use of ICT cannot yet be shown to enable learners to make greater progress than conventional instruction.
■ Nothing is known from quantitative research about the benefits of workplace basic skills provision.
■ The amount of instructional time needed for learners to make educationally significant progress is not securely known, but appears to be more than many spend in provision in a year.
■ No existing assessment instruments are fully suitable for use in NRDC's research.

A scoping review

1. Background

1.1 Adult literacy and numeracy

The publication of the Moser Report A Fresh Start (GB. DfEE, 1999) drew attention to the large number of adults in England and Wales with poor basic skills and gave this field greater prominence than ever before. An emphasis on related social factors was a feature of the report; it was specifically noted that there was much evidence connecting poor basic skills with issues such as unemployment and criminality. The government took the challenge of this situation very seriously and in November 2001 established the Adult Basic Skills Strategy Unit (ABSSU) within the (then) Department for Education and Employment (DfEE), now the Department for Education and Skills (DfES), to spearhead the work of tackling the problem. It also published a White Paper, **Skills for Life**, setting out its strategy (GB. DfEE, 2001). This included the target that by 2004 a total of 750,000 adults would significantly improve their basic skills, as a result of a concerted approach to provision, teaching and learning.

Skills for Life also foreshadowed the establishment of a National Research and Development Centre for Adult Literacy and Numeracy that would contribute to the improvement of basic skills teaching and learning. A consortium led by the University of London Institute of Education and including the Universities of Lancaster, Nottingham and Sheffield and several practitioner groups and national organisations as partners won the contract for the Centre, which formally came into existence in February 2002.

One of the main strands of the Centre's work focuses on best practice (pedagogy) and literacy and numeracy are two of the principal domains within this strand. In the early stages, much of the work within this strand consists of reviews of the research literature and the projects reported here (B1.1 and B1.5) consist of four of these reviews. Three of the reviews differ from the rest of NRDC's reviews in being a series of scoping and systematic reviews, in the specialised sense of those terms; that is, they are based on a particular, focused methodology, which attempts to cover all available published and 'unpublished' evidence on the relevant topic meeting certain rigorous criteria.

This series of scoping and systematic reviews has been designed to remedy the current lack of systematic reviews of all the randomised controlled trials (RCTs) and controlled trials (CTs) in the area of adult literacy and numeracy. Such reviews meet two aims: firstly, to assess whether any rigorous evaluations have been undertaken of interventions that could be implemented; and secondly, to inform the research agenda for future RCTs or CTs that may be used to evaluate interventions in adult literacy and numeracy.

1.2 Systematic reviews

There is an increasingly high profile of 'evidence-based policy' in education and educational research (Constable and Coe, 2000). A key element in the development of evidence-based education is the use of systematic review methodology. A systematic review differs from a traditional narrative review in that it seeks to identify all the available evidence with respect to a given theme. This evidence is collected, screened for quality and synthesised into an overall summary of the research in the field. Because all the evidence pertaining to a given topic is included in the systematic review, with rejected evidence catalogued and the reasons for

rejection made explicit, the resulting findings are often more objective than those of a non-systematic review. This is because a non-systematic review may include only evidence that reflects the biases of the author. The methodology of the systematic review aims to limit bias, to test hypotheses and to summarise the results of existing studies (Petticrew, 2001); the methods adopted are made explicit and transparent. Because of their perceived credibility, systematic reviews are often used to inform practice, scholarship and to provide a credible evidence base to support policy-making.

1.3 Why focus on randomised controlled trials and controlled trials?

Randomised controlled trials (RCTs)
The most rigorous quantitative study design for evaluating whether or not an intervention based on a causal question is effective is the RCT. This has been recognised in the field of social policy innovations for some time (Cook and Campbell, 1979; Cook, 2002). Within the wider research community, the superiority of the RCT over any alternative research design for making inferences about the effectiveness of interventions that examine causal relationships has been demonstrated (Cook, 2002). A synthesis of 74 meta-analyses of psychological, educational and behavioural interventions revealed that, although in these fields the average effect sizes did not differ between 'true' and 'quasi' experiments, the standard deviation was much larger for the non-randomised experiments (Lipsey and Wilson, 1993; Cook, 2002). In addition, one-group pre- and post-test designs for assessing the effectiveness of interventions can over-estimate the size of the effects by up to 61 per cent compared with evaluations using designs that employ a control or comparison group (Lipsey and Wilson, 1993).

Random assignment is used to exclude the possibility that any observed differences in outcome between the intervention and control groups are due to pre-existing differences. Randomised trials protect against both bias in attributing causal status to an independent variable and imprecision due to extraneous variation. Without random assignment it is difficult to justify inferences about cause and effect, because of the possibility of selection bias (Cook, 2002).

Controlled trials (CTs)
Quasi-experimentation is often the preferred methodology adopted by educational researchers. In questions of effectiveness, controlled trials are preferable to uncontrolled studies, such as before and after designs or post-test only methods. Controlled trials can reduce biases, such as temporal trends or regression to the mean effects, which pre-test/post-test designs are unable to do. Often controlled studies are used rather than the more rigorous randomised controlled studies for reasons of ease and economy.

1.4 Quality of RCTs and CTs included

Randomised controlled trials (RCTs)
Although RCTs are widely regarded as the 'gold standard' of effectiveness research, clearly their results are more reliable when the trials are of high quality. In the field of health care there has been considerable disquiet for many years over poorly conducted RCTs misinforming policy. Indeed, major medical journals now insist that reports of RCTs conform to the Consolidated Standards for Reporting Trials (CONSORT) guidelines (Begg et al. 1996). In this report we have used a modified version of the CONSORT guidance to appraise the quality of identified trials, rather than developing quality assurance instruments de novo and

in Appendix A we outline some key elements of the RCT method, so that our comments about the quality of included trials in the results section are set within a methodological context.

Controlled trials (CTs)
We used a further modification of the CONSORT guidelines to appraise the quality of the controlled trials. We removed those sections referring to description of randomisation procedures.

2. Objectives

The primary objectives of this scoping review were

- to search for, locate and synthesise all the RCTs and CTs intended to evaluate interventions in adult literacy and/or numeracy and published between 1980 and 2002;
- to establish what conclusions, if any, could be drawn from the RCTs and CTs about the effectiveness of teaching; and
- to investigate whether or not there was any difference in the pooled effect size of all the RCTs compared with the pooled effect size of all the CTs (where effect sizes could be calculated).

3. Methods

Studies were systematically searched for and located. They were then screened for inclusion in the scoping review using pre-established inclusion and exclusion criteria. At this screening stage of the review two researchers worked independently; they then conferred and agreed on all decisions. Those papers that appeared relevant from the abstracts were sent for; those that contained randomised or non-randomised controlled trials, or reviews of such trials, were read in full. At this second screening stage any papers not found to be relevant were excluded. The nine RCTs and 34 CTs included were data extracted for the scoping review. The findings were then synthesised in mapping tables. Effect sizes were calculated by the reviewers using the means and standard deviations of the post-test or the mean gain scores and standard deviations where available. Because some of the authors adjusted for lack of baseline equivalence by including the pre-test scores as a co-variate in the analysis some of the effect sizes calculated by the reviewers are different from those calculated by the authors.

Evidence from meta-analyses in educational research suggests that the effect sizes of RCTs and CTs are very similar, with a greater standard deviation in the controlled trials (Lipsey and Wilson, 1993). Therefore we calculated the effect sizes of the RCTs and compared them with the effect sizes of the CTs.

The rest of the methodology of the review is given in full in Appendix C.

4. Results

4.1 Identification of studies
A total of 4,555 potentially relevant papers were identified using the electronic and hand searches as described in Appendix D.

The 4,555 papers identified were double screened, using the inclusion/exclusion criteria

established in the protocol. Of these, 4,387 papers were excluded at this first stage and 168 papers were sent for. Within this subset, 19 papers were unavailable or were not received. All papers received were re-screened on the basis of a reading of the full paper. At this (second) stage, 90 papers were excluded, leaving a total of 59 papers to be included in the review. The titles and abstracts of all the papers in the database are available as an EndNote library on request (ar31@york.ac.uk), together with the reasons for their inclusion or exclusion. The 59 included papers were coded and the database was fully annotated with the codes (by Alison Robinson). 10 reviews were identified (and included) through the electronic searches and through a contact. One RCT was identified from a hand search of the bibliographies of these included reviews.

Table 1 presents the origin, by database or other method of retrieval, of all the 59 studies included in the scoping review.

Table 1: Origin of all included studies

	Found	**Included**
ERIC	2628	40
PsycINFO	971	3
CJA	736	5
SSCI	15	2
C2-SPECTR	8	1
SIGLE	172	1
Website	11	2
Handsearch	13	4
Contact	1	1
TOTAL	**4555**	**59**

4.2 Quality assurance

Screening
The electronic search strategy was very sensitive, but not very specific. Because of this relatively few studies were included and a large number of studies were excluded. Total agreement of the four main databases was 4,405 out of 4,524 (85 included and 4,320 excluded) or 97.37 per cent. Therefore, disagreement occurred in only 2.63 per cent of the total number of screened studies (119 studies). The two reviewers discussed these 119 studies and came to an agreement about whether or not to include them. The inter-rater reliability score between the two reviewers undertaking initial screening of titles and abstracts was 0.57 (moderate), because of the large size of the database relative to the number of included studies. The actual percentage disagreement was very small and not thought to be worrying. Full agreement between reviewers on exclusion at second stage was established (i.e. there was 100 per cent agreement to include and exclude).

Data extraction
Data were double entered onto pre-established data extraction tables by two reviewers working independently. In cases where there was initial disagreement about data extraction or quality appraisal this was discussed and resolved. Agreement was generally excellent. Full

agreement on calculation of effect sizes was established (i.e. 100 per cent agreement).

4.3 Descriptive map of all included studies

Table 2 presents the descriptive map of all the included studies. Of the 59 studies that were included in the scoping review, 12 papers were included that reported on nine RCTs. These RCTs were data extracted, quality appraised and summarised for the systematic review of RCTs. A further three papers reporting RCTs were found that included no data, or incomplete data, on adult literacy and numeracy. Some of these were preliminary studies; others presented data on other outcomes. Ten reviews were included, and 34 controlled trials. Nine of the CTs included sufficient data in order for the effect sizes to be calculated. These nine CTs were data extracted, quality appraised and summarised for the systematic review of CTs.

Table 2: Mapping of relevant RCTs, CTs and reviews

RCTs	12 papers containing 9 trials
RCTs (no results)	3 (Argento, 1980; Kemple et al. 1993; McCrossan, 1998)
CTs	34
Reviews	10

Table 3: Controlled trials

CTs with effect size	
Diem and Fairweather, 1980	Meyer et al. 1983
Dietrich, 1994	Nurss, 1989
Gretes and Green, 1994	Roberts et al. 1994
Lavery et al. 1998	Schrader, 1984
Maclay and Askov, 1988	

CTs with no effect size (incomplete data)	
Askov et al. 1986	Macmurdo, 1988
Broughton, 1994	Morrow et al. 1993
Broussard, 1983	Rio Salado Community College, 1991
Burtoff, 1985	Smith and Dalheim, 1990
Culclasure, 1982	Thuy, 1992
Dirx and Crawford, 1993	Wadsworth and Frazier, 1982
Indiana OIC, 1990	Washington DoE, 1991
Irby et al. 1992	Wilson, 1992
Lehigh Community College, 1993	Wisher and O'Hara, 1981

CTs with no results	
Conti, 1980	Newsom, 1980
Cox, 1981	Ramsey, 1983
Enriquez, 1990	Tewksbury, 1994
Johnston, 1980	

Table 4: Reviews

Adult Literacy and Basic Skills Unit, 1994
Beder, 1999
Bouffard, 2000
Jancic, 1998
Kruidenier, 2002
Kulik, 1986
Rachal, 1995
Scanlon et al. 1998
Sticht, 1995
Wilson, 1999

Altogether there were 12 RCTs and 34 CTs. Of these 46 trials, 36 (nine RCTs and 27 CTs) contained data (three RCTs and seven CTs contained no data). The remainder of this scoping review focuses on the 36 included trials.

Table 5: Country of origin

RCTs	
USA	8
UK	1
CTs	
USA	26
New Zealand	1

Of all the 36 included trials with data, 34 were undertaken in the USA and only one (RCT) was undertaken in the UK and one (CT) in New Zealand.

Table 6: Adult basic education focus by study design

Study Design	Literacy only	Numeracy only	Literacy and Numeracy	Total
RCT	5	2	2 (1)	9 (1)
CT	19	0	8	27
TOTAL	24	2	10 (1)	36 (1)

(Numbers in brackets represent studies conducted in the UK)

4.4 Studies with a literacy focus

Thirty-four of the adult basic education studies had a literacy focus, of which seven were RCTs and 27 were CTs (Table 6). Of those 34 studies, 10 also examined numeracy. Literacy was assessed by a variety of outcome measures (grade or literacy test scores), with each study measuring at least one outcome.

In most of the studies the intervention consisted of CAI – 18 studies. Other interventions included small group instruction versus one-to-one instruction; a phonological skills

approach versus a traditional meta-cognitive approach; various family literacy initiatives, etc. The comparison groups received traditional literacy instruction or programmes, with the exception of four studies that used a no-treatment control (Maclay and Askov, 1988; Schrader, 1984; Thuy, 1992; and St. Pierre et al. 1993).

Five studies reported a positive effect for the intervention compared with the control group (the 'diagnostic prescriptive approach' to reading and the 'reciprocal teaching programme', both of which were RCTs, and three CTs) and one study reported a positive effect for the control group, i.e. a negative effect for the intervention. Ten studies showed no difference. The results of the remaining 18 studies were unclear because it was not possible for the reviewers to calculate an effect size. These results, however, should be interpreted with caution, as the quality of the CTs with insufficient data was not assessed. Of the 34 studies, 33 were undertaken in the USA and one was undertaken in the UK.

4.5 Studies with a numeracy focus

12 of the adult basic education studies had a numeracy focus, of which four were RCTs and eight were CTs (Table 6). Of those 12 studies, ten also examined literacy. Numeracy was assessed by a variety of outcome measures (grade or numeracy test scores), with each study measuring at least one outcome.

In most of the studies the intervention consisted of CAI. Other interventions included a modified comprehension learning strategy, 'Even Start' interventions and a family literacy programme with a numeracy element. The comparison groups received traditional adult basic education classes, with the exception of two studies that also had a no-teaching group (Nicol and Anderson, (2000) had a CAI group, a conventional teaching group and a no-teaching group that consisted of structured activities but did not directly teach numeracy; St. Pierre et al. (1993) used a control group that had not participated in 'Even Start').

One study reported a positive effect for the intervention compared with the control group (the modified comprehension learning strategy, which was an RCT), while six showed no difference (two RCTs and three CTs that were all CAI interventions, and one other RCT). The results of the remaining five studies were unclear (four CTs of CAI and one CT of a family literacy programme) because it was not possible for the reviewers to calculate an effect size. These results, however, should be interpreted with caution, as the quality of the CTs with insufficient data was not assessed.

Of the 12 studies, ten were undertaken in the USA, one in New Zealand and one in the UK. In the RCT that took place in the UK (Nicol and Anderson, 2000) the effect of CAI on the numeracy skills of adults with 'mild learning disabilities' was assessed. The authors reported that there was no difference between the CAI group and the traditional teaching group.

Tables 7 and 8 summarise the studies according to topic focus (literacy or numeracy), and include information on type of study, intervention, outcome and effect.

Table 7: Summary of studies with a literacy focus

Key:

* = interventions and outcomes in both literacy and numeracy;

** 95% confidence intervals;

*** reviewers unable to calculate effect sizes.

Author, year, country	Type of study	Intervention	Outcome	Effect of intervention
*Batchelder and Rachal, 2000a (USA)	RCT	GED (US General Educational Development) instructional material on computers plus traditional instruction in English, maths, history and science vs traditional instruction in English, maths, history and science.	Comprehensive Adult Student Assessment System (CASAS) reading post-test.	■ EFFECT No difference ■ EFFECT SIZE 0.26 (unadjusted as reported by authors)
Bean and Wilson, 1989 (USA)	RCT	Closed caption television with or without instruction as a reading medium for sight vocabulary development vs traditional print instruction.	Sight vocabulary recognition test.	■ EFFECT No difference ■ EFFECT SIZE 0.77 (-0.24 to 1.74)**
Cheek and Lindsey, 1994 (USA)	RCT	Diagnostic prescriptive approach to reading comprehension vs traditional programmed approach.	Stanford diagnostic reading test (word identification and reading comprehension).	■ EFFECT Positive effect for intervention ■ EFFECT SIZE 0.46 (-0.02 to 0.92)**
Martinson and Friedlander, 1994 (USA)	RCT	Access to California's state-wide greater avenues for independence programme vs no access to programme.	The test of applied literacy skills (TALS)	■ EFFECT No difference ■ EFFECT SIZE 0.019 (as reported by authors)
McKane and Green, 1996 (USA)	RCT	Access to computer-assisted reading instruction plus traditional reading instruction vs traditional reading instruction alone.	Test for Adult Basic Education (TABE) post-test.	■ EFFECT No difference ■ EFFECT SIZE 02(-0.21 to 0.61)**
Rich and Shepherd, 1993 (USA)	RCT	Reciprocal teaching (text comprehension) programme (self-questioning and summarising, or self-questioning or summarising) vs experimental materials without instruction.	Multiple-choice short-answer comprehension test and free recall comprehension test.	■ EFFECT Positive effect for intervention (self questioning and summarising and self-questioning but not summarising). ■ EFFECT SIZE 1.22 and 0.92(as reported by authors)
*St. Pierre et al. 1993 (USA)	RCT	Variety of activities including adult education, adult basic education, adult secondary education, English as a second language and preparation to attain a GED certificate.	CASAS reading	■ EFFECT No difference ■ EFFECT SIZE 0.70 (0.05 s.d.) $p \leftarrow 0.05$, as reported by authors; 0.29, ns, as calculated by reviewers
Askov et al. 1986 (USA)	CT	Computer-assisted instruction in literacy vs small group and tutorial instruction in traditional materials for literacy vs no exposure to CAI.	Bader reading and language inventory	■ EFFECT Unclear ■ EFFECT SIZE ***
*Broughton, 1994 (USA)	CT	Use of computers in reading instruction vs standard instruction.	TABE	■ EFFECT Unclear ■ EFFECT SIZE ***
*Broussard, 1993 (USA)	CT	Computer-assisted instruction using adult basic curriculum vs traditional adult basic education curriculum.	Grade equivalence from California Achievement Test.	■ EFFECT Unclear ■ EFFECT SIZE ***

Author, year, country	Type of study	Intervention	Outcome	Effect of intervention
Burtoff, 1985 (USA)	CT	Haitian Creole literacy instruction followed by ESL vs ESL instruction only.	ESL proficiency, basic English skills.	■ EFFECT Unclear ■ EFFECT SIZE ***
Culclasure, 1982 (USA)	CT	Basic literacy instruction (reading comprehension) using computers vs basic literacy instruction using traditional classroom teaching approaches.	California achievement test	■ EFFECT Unclear ■ EFFECT SIZE ***
*Diem and Fairweather, 1980 (USA)	CT	PLATO (Program Logic for Automated Training Operations) interactive computer graphics terminal model for instruction in literacy and numeracy vs traditional instruction in literacy and numeracy.	Adult basic learning examination level 2	■ EFFECT No Difference ■ EFFECT SIZE Vocabulary = 0.56 (-0.17 to 1.29) ** Reading = 0.16 (-0.56 to 0.88) Spelling = -0.64 (-1.37 to 0.10) ** Arithmetic = 0.02 (-0.70 to 0.73) **
Dietrich, 1994 (USA)	CT	Phonological skills approach to reading vs traditional meta-cognitive approach to reading.	Lindamood Auditory Conceptualisation test, Woodcock Johnson achievement test (word identification and word attack).	■ EFFECT No difference ■ EFFECT SIZE LAC = 0.85 (-0.06 to 1.74) ** Word attack = -0.63 (-1.51 to 0.25) **
Dirk and Crawford, 1993 (USA)	CT	Contextual learning approach to literacy vs traditional individual.	Reading levels	■ EFFECT Unclear ■ EFFECT SIZE ***
Gretes and Green, 1994 (USA)	CT	Computer-delivered reading instruction vs traditional reading instruction.	TABE total reading	■ EFFECT Large positive effect for intervention ■ EFFECT SIZE 0.97 (0.782 to 1.157) **
*Indiana OIC of American State Council, 1990 (USA)	CT	Computer-assisted teaching vs traditional classroom teaching.	ABLE test	■ EFFECT Unclear ■ EFFECT SIZE ***
*Irby et al. 1992 (USA)	CT	Family literacy project vs GED classes only	TABE	■ EFFECT Unclear ■ EFFECT SIZE ***
*Lavery et al. 1998 (New Zealand)	CT	Computer-assisted instruction in basic literacy and numeracy vs traditional instruction in basic literacy and numeracy	Burt word reading test Neale analysis of reading ability Keymath revised test	■ EFFECT No difference ■ EFFECT SIZE Burt 1.24 (-0.04 to 2.46) ** Neale Accuracy 0.48 (-0.68 to 1.62) ** Neale comprehension –0.64 (-1.79 to 0.54) ** Math total –0.77 (-1.93 to 0.43) **
Lehigh County Community College, 1993 (USA)	CT	GED instruction emphasising functional and workplace context vs traditional GED instruction.	The test of applied literacy skills (TALS)	■ EFFECT Unclear ■ EFFECT SIZE ***
Maclay and Askove, 1988 (USA)	CT	Computer-assisted instruction in literacy vs no teaching.	Slosson oral reading test.	■ EFFECT Positive effect for intervention ■ EFFECT SIZE 0.82 (0.567 to 1.073) **
Macmurdo, 1988 (USA)	CT	Reading instruction using volunteer tutors and CAI vs reading instruction using volunteer tutors without CAI	Grade levels	■ EFFECT Unclear ■ EFFECT SIZE ***

Author, year, country	Type of study	Intervention	Outcome	Effect of intervention
Morrow et al. 1993 (USA)	CT	Small group instruction in reading vs traditional one-to-one instruction in reading.	Reading comprehension test and word recognition test.	■ EFFECT Unclear ■ EFFECT SIZE ***
Nurss, 1989 (USA)	CT	Computer-assisted instruction for literacy vs traditional non-computerised programme for literacy.	TABE	■ EFFECT Positive effect for control ■ EFFECT SIZE -1.63 (-2.36 to –0.89) **
Rio Salado Community College, 1991 (USA)	CT	Home-based CAI in addition to traditional ABE classes vs traditional ABE classes.	TABE	■ EFFECT Unclear ■ EFFECT SIZE ***
Roberts et al. 1994 (USA)	CT	Community building group process and SRA reading programme vs SRA reading programme.	Gates-MacGinitie reading test.	■ EFFECT Positive effect for intervention ■ EFFECT SIZE GE = 1.1 (0.57 to 1.62) ** ESS = 1.0 (0.48 to 1.51) **
Schrader, 1984 (USA)	CT	A developmental reading programme vs no treatment control.	Grade point average	■ EFFECT No difference ■ EFFECT SIZE -0.063 (-0.66 to 0.53) **
Smith and Dalheim, 1990 (USA)	CT	Structured phonics approach to reading with or without equipment vs language experience approach	Oral reading and word attack Oral and silent reading comprehension	■ EFFECT Unclear ■ EFFECT SIZE ***
Thuy, 1992 (USA)	CT	Computer-assisted instruction in ESL and family literacy curriculum vs printed materials only	Vocabulary and comprehension questions	■ EFFECT Unclear ■ EFFECT SIZE ***
*Wadsworth and Frazier, 1982 (USA)	CT	Computer-assisted instruction for literacy and numeracy for less than 20 hours vs computer assisted instruction for literacy and numeracy for more than 20 hours	Grade level achievement	■ EFFECT Unclear ■ EFFECT SIZE ***
Washington (US Dept of Education), 1991 (USA)	CT	Computer-assisted instruction plus traditional literacy classes vs traditional literacy classes	Assessment test developed by college	■ EFFECT Unclear ■ EFFECT SIZE ***
*Wilson, 1992 (USA)	CT	Computer-assisted instruction for literacy and numeracy vs traditional approaches for literacy and numeracy	TABE	■ EFFECT Unclear ■ EFFECT SIZE ***
Wisher and O'Hara, 1981 (USA)	CT	Academic remedial training through CAI vs academic remedial training through classroom instruction.	Gates-MacGinitie reading test, navy recruit reading test	■ EFFECT Unclear ■ EFFECT SIZE ***

Table 8: Summary of studies with a numeracy focus

Key:

* = interventions and outcomes in both literacy and numeracy;

** 95% confidence intervals;

*** reviewers unable to calculate effect sizes.

Author, year, country	Type of study	Intervention	Outcome	Effect of intervention
*Batchelder and Rachal, 2000a, USA	RCT	GED instructional material on computers plus traditional instruction in English, maths, history and science vs traditional instruction in English, maths, history and science.	CASAS	■ EFFECT No difference ■ EFFECT SIZE 0.16
Nicol and Anderson, 2000, UK	RCT	Computer-assisted teaching for numeracy vs conventional teaching for numeracy.	Two psychometric tests of numeracy: The basic number diagnostic test, Gillham, 1980 The basic number screening test, Gillham and Hesse, 1976	■ EFFECT No difference ■ EFFECT SIZE Post-test1 0.46 (-0.55 to 1.44) ** Post-test2 0.53 (-0.47 to 1.52) **
Schrum, 1985, USA	RCT	Modified comprehension learning strategy system vs conventional methods of teaching with material presented in conventional lecture method (blackboard).	Mastery of basic arithmetical computations as measured by wide range achievement test, arithmetic, Level II (1978)	■ EFFECT Positive effect for intervention ■ EFFECT SIZE Post-test1 0.64 (0.17 to 1.11) ** Post-test 2 0.77 (0.29 to 1.25) **
*St. Pierre et al. 1993, USA	RCT	Variety of activities including adult education, adult basic education, adult secondary education, English as a second language and preparation to attain a GED certificate.	CASAS	■ EFFECT No difference ■ EFFECT SIZE 0.70 (0.05 s.d.) p ← 0.05, as reported by authors; 0.29, ns, as calculated by reviewers
*Broughton, 1994, USA	CT	Use of computers in reading instruction vs standard instruction.	TABE	■ EFFECT Unclear ■ EFFECT SIZE ***
*Broussard, 1983, USA	CT	Computer-assisted instruction using adult basic curriculum vs traditional adult basic education curriculum.	Grade equivalence from California achievement test	■ EFFECT Unclear ■ EFFECT SIZE ***
*Diem and Fairweather, 1980, USA	CT	PLATO interactive computer graphics terminal model for instruction in literacy and numeracy vs traditional instruction in literacy and numeracy.	Arithmetic computation and arithmetic problem as measured by adult basic learning examination (ABLE) Level II	■ EFFECT No difference ■ EFFECT SIZE Arithmetic 0.02 (-0.70 to 0.73) **
*Irby et al. 1992, USA	CT	Family literacy project vs GED classes only	TABE	■ EFFECT Unclear ■ EFFECT SIZE ***
*Lavery et al. 1998, New Zealand	CT	Computer-assisted instruction in basic literacy and numeracy vs traditional instruction in basic literacy and numeracy	Basic maths processes and mathematics problem solving as measured by KeyMath Revised Test	■ EFFECT No difference ■ EFFECT SIZE Math total –0.77 (-1.93 to 0.43) **

Author, year, country	Type of study	Intervention	Outcome	Effect of intervention
*Meyer et al. 1983, USA	CT	Computer-managed programme for reading, maths and language instruction vs traditional programme for reading, maths and language instruction.	TABE	■ EFFECT No difference ■ EFFECT SIZE Math 0.2466 (-0.071 to 0.5636) **
*Wadsworth and Frazier, 1982, USA	CT	Computer-assisted instruction for literacy and numeracy for less than 20 hours vs computer assisted instruction for literacy and numeracy for more than 20 hours	Grade level achievement	■ EFFECT Unclear ■ EFFECT SIZE ***
*Wilson, 1992, USA	CT	Computer-assisted instruction for literacy and numeracy vs traditional approaches for literacy and numeracy	TABE Canadian Adult Achievement Test	■ EFFECT Unclear ■ EFFECT SIZE ***

A comparison of the effect sizes of the RCTs of the CTs

The graphs below show a comparison of the effect sizes from the RCTs and the CTs. The graphs show that generally there seem to be similar effect sizes from both study types. The average effect size for the CTs is 0.33 (95% CI −0.13 to 0.79), whilst for the RCTs the average effect size was 0.39 (95% CI 0.14 to 0.64) [estimated using a random effects model meta-analysis].

Funnel plot of Effect Size against Sample Size
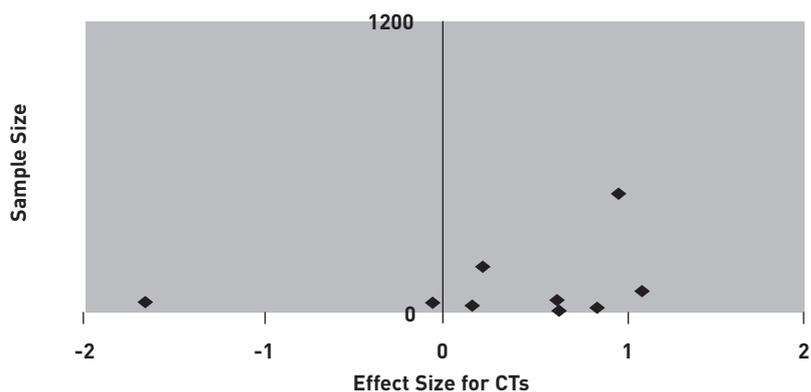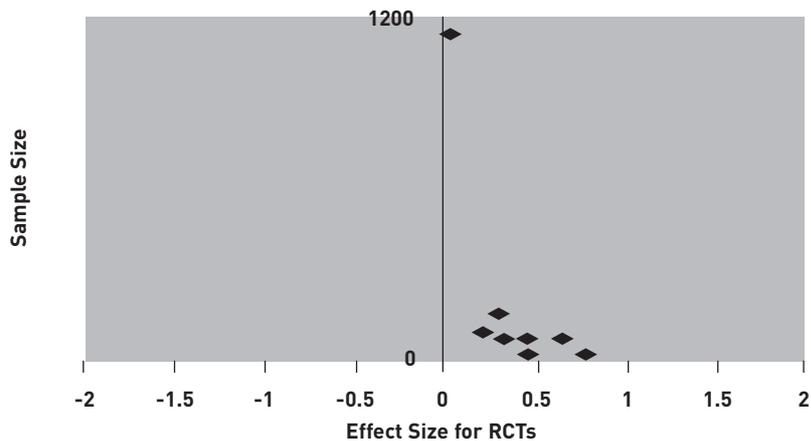
4.6 Discussion

Out of a total of 46 trials (12 RCTs and 34 CTs) included in this scoping review only 36 contained data (9 RCTs and 27 CTs), and of those 36 trials only 18 (9 RCTs and 9 CTs) either reported effect sizes or contained sufficient data for the reviewers to calculate effect sizes. Only six studies out of a total of 36 trials showed a statistically significant positive outcome for the intervention as determined from the effect size. Three of these studies were RCTs. These were: Cheek and Lindsey, 1994 (literacy), Rich and Shepherd, 1993 (literacy) and Schrum, 1985 (numeracy). These studies are examined in detail later in this report (the systematic review of randomised controlled trials). Three of these studies were CTs. These were Gretes and Green, 1994 (literacy); Maclay and Askov, 1988 (literacy) and Roberts et al. 1994 (literacy). These studies are examined in more detail later in this report (the systematic review of controlled trials).

4.8 Conclusions

This overview of all the experimental research in the fields of adult literacy and numeracy since 1980 has revealed only six studies out of a total of 36 trials showing a statistically significant positive outcome for the intervention, i.e. a promising intervention or pedagogy that could be used to inform policy or practice.

A systematic review of randomised controlled trials (RCTs)

1. Introduction

1.1 Background

Ideally, before policy interventions are put into place they should be subjected to a rigorous evaluation using a large well-designed randomised controlled trial (RCT). Failure to evaluate before implementation can lead to policies that are sub-optimal. To aid 'evidence-based' policy making what is first required is, a systematic review of all the available rigorous evidence. A recent overview of the literature in this field by Brooks et al. (2001a) concluded that few intervention studies had been carried out in the area of adult literacy and that no systematic review of intervention studies on adult literacy and numeracy had been undertaken. In particular, they found no RCTs. They did find and review a number of national and other large-scale studies of progress in adult literacy and numeracy, drawing mainly on Beder (1999), Brooks et al. (2001a) and Sticht and Armstrong (1994). The general conclusion from the studies was that progress was both slow and modest.

1.2 The research questions

The research question for the systematic review of RCTs was: What evidence is there of the effectiveness of interventions or pedagogies for increasing literacy or numeracy in adults? Two further focuses for the search were:

- studies that include a follow-up assessment of learners at some point after an intervention (in order to judge the extent to which gains while in provision are sustained afterwards); and
- wider benefits of learning.

1.3 Objectives

The primary objectives of this systematic review were:

- to search for, locate, synthesise and quality appraise all the RCTs intended to evaluate interventions in adult literacy and/or numeracy, published between 1980 and 2002; and
- from those RCTs to establish what conclusions, if any, could be drawn about the effectiveness of teaching.

Other objectives were:

- to analyse what the same studies had to say about the wider benefits to the adult learner of participating in adult literacy and numeracy interventions;
- to draw methodological conclusions about the quality of the research covered; and
- to draw up a proposal for a well-designed RCT on a key question in the field.

2. Methods

A protocol (see Appendix B) was developed for the systematic review of RCTs in order to establish: the research question; the scope and limitations of the review; the methods for

conducting the review; the inclusion and exclusion criteria; and the procedure for the extraction of data and quality appraisal. The protocol was developed by two members of the review team with assistance from members of the advisory panel. Methods for undertaking the scoping review and the systematic review of CTs were based on this protocol.

The nine RCTs were independently double data extracted and quality appraised for the systematic review. Five of the RCTs were entered into two meta-analyses. Information on wider benefits of learning was qualitatively analysed from the included RCTs.

The rest of the methodology of the review is given in full in Appendix C.

3. Results

3.1 Description of included RCTs

Of the nine included RCTs, five evaluated interventions in literacy, two evaluated interventions in numeracy and two evaluated interventions in both (Table 1).

Table 1:   Study topic of included RCTs

Literacy	Numeracy	Literacy and Numeracy
Bean and Wilson, 1989 (Reading)	Nicol and Anderson, 2000	Batchelder and Rachal, 2000 a and b
Cheek and Lindsey, 1994 (Reading and Comprehension)	Schrum, 1985	St. Pierre et al. 1993
Martinson and Friedlander,1994		
McKane and Greene, 1996 (Reading)		
Rich and Shepherd, 1993 (Reading and Comprehension)		

3.2 Methodological quality of studies

All of the trials included were of a high quality in the sense that they used an appropriate study type for questions of effectiveness: the RCT. Table 2 shows whether or not the studies fulfilled our modified CONSORT checklist of quality criteria. As the table shows, the quality of the trials was variable, with no trial concealing random allocation and only one trial using intention to teach analysis. The study that appeared to be the most robust, methodologically, was Rich and Shepherd (1993). This was the only study that used blinded assessment at follow-up and appeared to be the only one to use intention to teach analysis. Detailed information on the RCTs can be found in Appendix E; and in Appendix G we consider the quality of each trial in depth.

Table 2: Quality of included RCTs

Y = Yes N = No N/S = not stated

U = unclear ind = individual

	Batchelder and Rachal 2000	Bean and Wilson, 1989	Cheek and Lindsey, 1994	Martinson and Friedlander, 1994	McKane and Greene,1996	Nicol and Anderson, 2000	Rich and Shepherd, 1993	Schrum, 1985	St. Pierre et al. 1993
Was the study population adequately described? (i.e. were the important characteristics of the randomised adults described e.g. age, gender?)	Y	Y	Y	N	N	Y	Y	Y	N
Was the minimum important difference described? (i.e. was the smallest educationally important effect size described?)	N/S	N/S	N/S	N/S	N/S	N/S	N/S	N/S	N/S
Was the target sample size adequately determined?	N/S	N/S	N/S	N/S	N/S	N/S	N/S	N/S	N/S
Was intention to treat analysis used? (i.e. were all adults who were randomised included in the follow-up and analysis?)	N	N	N	N	N	N/S	Y	N/S	N
Was the unit of randomisation described (i.e. individual adults or groups of adults)?	Y IND	Y IND	Y IND	Y IND	Y IND	Y IND	Y IND	Y IND	Y IND
Were the participants allocated using random number tables, coin flip, computer generation?	Y	N/S	N/S	N/S	N/S	N/S	N/S	Y	N/S
Was the randomisation process concealed from the investigators? (i.e. were the researchers who were recruiting adults to the trial blind to the adult's allocation until after that adult had been included in the trial?)	N/S	N/S	N/S	N/S	N/S	N/S	N/S	N/S	U
Were follow-up measures administered blind? (i.e. were the researchers who administered the outcome measures blind to treatment allocation?)	N/S	N/S	N/S	N/S	N/S	N/S	Y	N/S	Y
Was estimated effect on primary and secondary outcome measures stated?	Y	Y	Y	U	Y	Y	Y	Y	Y
Was precision of effect size estimated (confidence intervals)?	N/S	N	N	N/S	N/S	N	N/S	N/S	N
Were summary data presented in sufficient detail to permit alternative analyses or replication?	Y	Y	Y	U	Y	Y	Y	N	Y
Was the discussion of the study findings consistent with the data?	Y	N	Y	N	Y	N	Y	Y	Y

3.3 Publication bias

Funnel plot of Effect Size against Sample Size for RCTs
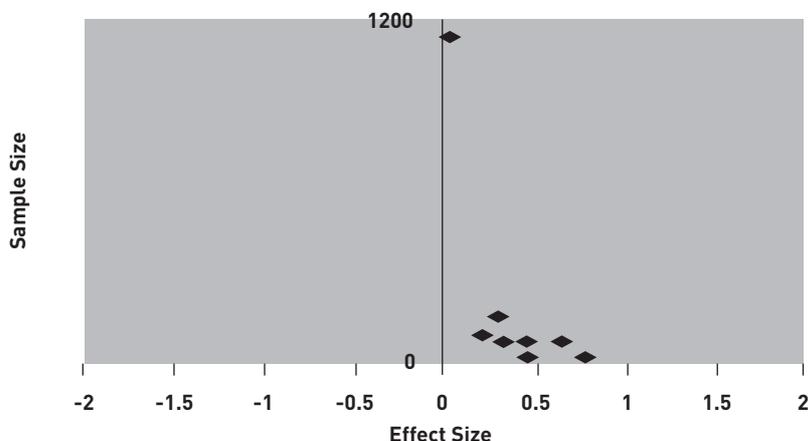
The funnel plot (Figure 1) indicates that there is probably publication bias. All the effect sizes of the rest of the studies fall to the right of the effect size of the largest, with none falling to the left. If the studies included in this review were either all the RCTs that had been conducted or a true random sample, then we would expect some of the smaller studies to fall to the left of the largest trial. Because this is not happening this suggests that some of the smaller, negative studies (i.e. those showing adult literacy and numeracy interventions were harmful) were not published either as a journal article or within the grey literature to which we had access. We can compare the above plot with a hypothetical funnel plot below (Figure 2), which shows a group of trials with no publication bias.

Hypothetical Funnel Plot showing little Publication Bias
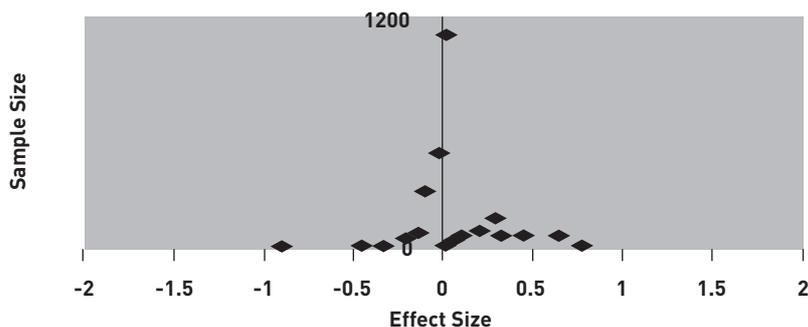
This plot shows little or no publication bias as there are trials with effect sizes falling in the left-hand side of the graph, whereas Figure 1 shows a complete absence of trials on that side.

3.4 Meta-analysis of published and unpublished studies

We compared the mean effect sizes of the studies that have been published (that is, formally, e.g. in refereed journals) with those that are 'unpublished' (that is, published informally, e.g. as in-house reports or mimeographs, or so-called 'grey' literature). The six published studies show a pooled effect size of 0.49 (95% CI 0.17 to 0.82, p = 0.003), whereas the three 'unpublished' studies show a lower effect size of 0.26 (95% CI –0.07 to 0.59, p = 0.13). Clearly, the unpublished studies are published in the sense that their reports are obtainable and are in the public domain. Nevertheless, these reports tend to show an effect size approximately

half as large as the studies that are published in journals. Studies that have never been published in any form are likely to have even smaller effect sizes. These data coupled with the funnel plot indicate that there are probably significant numbers of 'missing' studies with either negative, null or very small positive effect sizes.

3.5 Study outcomes

Eight of the nine included RCTs were undertaken in the USA, whilst the remaining study was undertaken in Scotland. As this table shows, seven of the nine trials showed a positive effect for the interventions; in two instances this was statistically significant and in one instance the effect was close to statistical significance.

Table 3 also shows that the interventions and the study populations were very heterogeneous. This heterogeneity precludes a meta-analysis of all the studies, because interventions that may be effective in one setting, a non-prison population, for example, may not be as effective within another setting, for example a prison setting. Nevertheless, there were two subsets of studies that were sufficiently similar to justify combining them in meta-analyses.

3.6 Meta-analysis of adult learning programmes against no treatment

As a starting point we wished to assess whether adult literacy and numeracy teaching was beneficial against no teaching. This is important because clearly participants in such programmes had not become sufficiently literate or numerate through exposure to traditional childhood education. One might therefore hypothesise that for such adults any form of teaching in adulthood could again be ineffective.

Table 3 Description of included RCTs

Country	Population	Intervention	Effectiveness
Batchelder and Rachal, (2000 a and b), USA	Male inmates in maximum security prison	Use of CAI in literacy and numeracy instruction	Positive effect for literacy and numeracy (not statistically significant)
Bean and Wilson (1989), USA	Students attending adult literacy programme	Use of closed caption television as a reading medium for sight vocabulary development	Positive effect for intervention (not statistically significant)
Cheek and Lindsey (1994), USA	Economically and academically disadvantaged adults attending an urban vocational technical institute	Comparison of two methods of reading instruction (diagnostic, prescriptive and traditional programmes)	Positive effect for total comprehension (borderline statistical significance)
Martinson and Friedlander (1994), USA	Welfare recipients meeting GAIN (Greater Avenues for Independence Program) criteria for basic education	Mix of interventions (e.g. small group and whole class teaching, CAI, individualised programmes, peer-based instruction, published work book, teacher/student development materials)	Virtually no programme impact on basic skills levels on all counties as a group
McKane and Greene (1996), USA	Male and female inmates in maximum security prisons	Comparison between CAI and non-CAI for reading instruction	Small positive effect for intervention (not statistically significant)
Nicol and Anderson (2000), UK	Adults with 'mild learning disabilities'	Comparison between CAI and conventional teaching of numeracy	Positive effect for CAI but not statistically significant
Rich and Shepherd (1993), USA	Students enrolled in adult education programmes	Use of reciprocal teaching (text comprehension) programme for reading comprehension	Positive and statistically significant effects for all 3 interventions versus two control treatments
Schrum (1985), USA	Male inmates in prison	Comparison of a modified comprehension learning strategy system for teaching numeracy with conventional methods	Positive and statistically significant effects
St. Pierre et al. (1993), USA	Even Start Families, mixed gender, ethnicity and ages	Variety of activities that included adult education (e.g. ABE, ASE, ESL, GED)	Small positive effect but not statistically significant

In Figure 3 we pooled the three studies that compared teaching against a no-teaching control group (Bean and Wilson, 1989; Nicol and Anderson, 2000; Rich and Shepherd, 1993). As the figure shows, all three trials showed a positive effect of teaching either literacy or numeracy, with one trial being statistically significant. When the three studies were pooled in a meta-analysis, teaching was associated with a large (0.88) and statistically significant effect size. This suggests, therefore, that offering literacy or numeracy teaching to adults will improve their literacy or numeracy. Though apparently obvious, this finding has not been rigorously demonstrated before and it is therefore significant, especially given the null hypothesis (stated above) that for adults with previous poor educational experiences any form of teaching in adulthood might again be ineffective. However, this observed beneficial effect must be tempered with the previous observations that there is publication bias in these data. Had missing negative or null studies been included the observed effect might have been weaker.

Moreover this analysis does not tell us which method of teaching literacy or numeracy is most effective. We therefore inspected the included RCTs for any which might be sufficiently alike to support a meta-analysis of a particular form of intervention.

Figure 3: Meta-analysis of teaching adults literacy or numeracy against no teaching

Cochrane effect size plot (fixed effects)

pooled effect size = 0.8811 (95% CI = 0.5031 to 1.2591)

Does not Favour Teaching   Favours Teaching
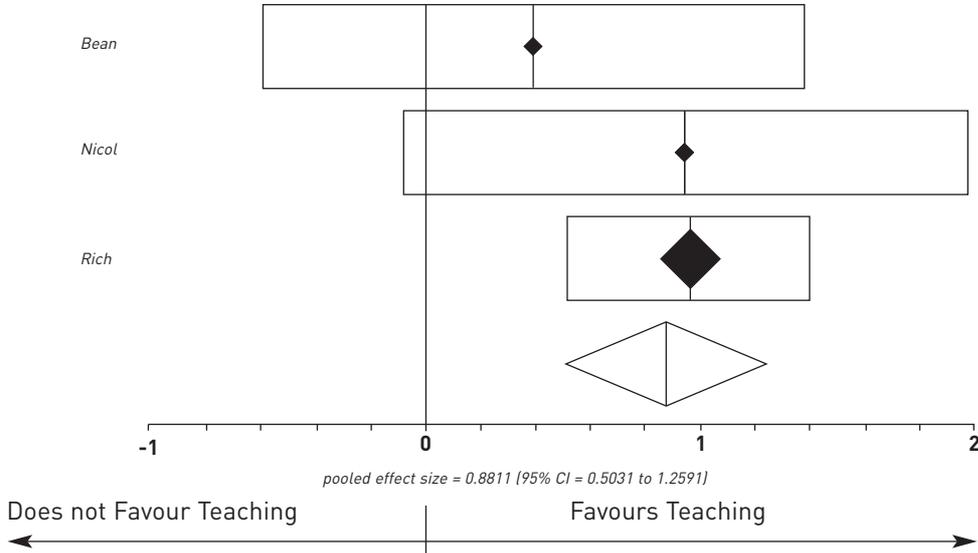
The size of the black diamond in each rectangular box is proportional to the sample size (i.e. the larger the diamond the bigger the sample size). The diamond at the bottom of the chart represents the summary pooled estimate of all the trials and the summary estimate of the 95 per cent confidence intervals.

3.7 Meta-analysis of CAI and incarcerated adults

The only area that we considered sufficiently homogeneous to undertake a meta-analysis of a single intervention was the use of CAI in adult literacy among imprisoned adults. Figure 4 shows that combining the two studies that looked at CAI in a prison setting (McKane and Greene, 1996; Batchelder and Rachal, 2000 a and b) gave a positive effect of about a fifth of a standard deviation; however, this effect was not statistically significant.

Figure 4: Meta-analysis of computer-assisted instruction (CAI) for adult literacy in prison settings

Cochrane effect size plot (fixed effects)
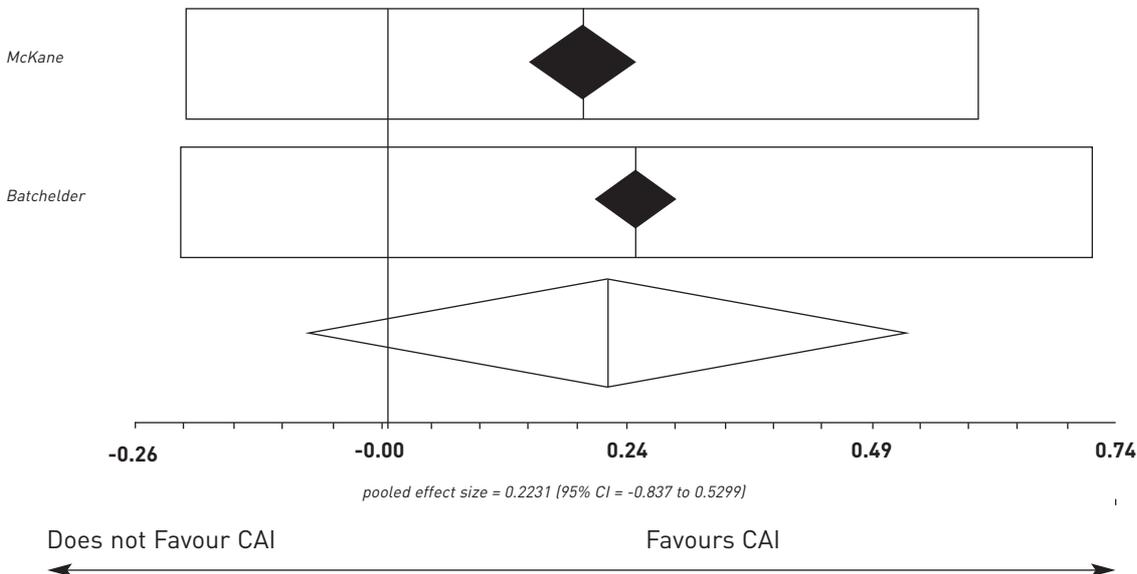
pooled effect size = 0.2231 (95% CI = -0.837 to 0.5299)

Does not Favour CAI   Favours CAI

A previous systematic review comparing computer-assisted and traditional approaches in adult reading achievement included 21 studies between 1979 and 1995:

> 'Despite considerable variability in design sophistication and clarity of reporting, ten of the 21 studies found no statistically significant differences between the two methods, five failed to indicate significance, two showed mixed results on differing assessment measures, three showed significance favoring CAI and one showed significance favoring traditional methods' (Rachal, 1995, p.239)

Although Rachal (1995) was also a systematic review, the dates for including studies differed from the dates for including studies in our systematic review, as did other inclusion and exclusion criteria (age, outcome measures, unpublished theses – we only included RCTs etc.). However, 11 of the 21 studies are included in our review.

3.8 Follow-up assessments
Two studies (Schrum, 1985; Nicol and Anderson, 2000) undertook two post-tests. We compared the pooled value of the first post-test with the second in order to see whether there was any evidence of a declining effect of the intervention. In the first pooled analysis (random effects model) the effect size was 0.60 (95 per cent CI 0.17 to 1.02), whilst the second post-tests showed a similar effect of 0.71 (95 per cent CI 0.28 to 1.15). This indicates that the benefit of the intervention in these two studies persisted.

3.9 Wider benefits of learning
Of the nine studies from which data extraction was completed, two (Cheek and Lindsey, 1994; McKane and Greene, 1996) contained no qualitative data. The remaining seven studies were fairly evenly divided amongst literacy and numeracy interventions (three literacy, two numeracy and two covering both). Table 4 summarises the findings on the wider benefits of these interventions and full analyses are given in Appendix H. The indicators listed are those that were found in each study to have had an impact, whether positive or negative. If an indicator is not listed it either had no significant impact or was never a consideration in the particular study. Our comments on the wider benefits findings are divided into content and quality.

With one significant exception (Batchelder and Rachal, 2000 a and b), the interventions were reported to have had positive wider benefits, especially in the numeracy studies and in one of the mixed studies (St. Pierre et al. 1993). It was not always clear which benefits were reported by participants and which by teachers, or to what proportion of participants the indicators applied, drawbacks which will be discussed in more detail when assessing quality.

A limited range of indicators of wider benefits was found in the studies, with the exception of St. Pierre et al. (1993). All but one study (Schrum, 1985) included comments on participant attitudes to the instructional methods used or to testing, all favourable except for Batchelder and Rachal (2000 a and b) and possibly Martinson and Friedlander (1994). Both of these studies involved non-voluntary participants. The former was a prison study using CAI, the latter involved welfare recipients who received a mixture of traditional instruction and CAI.

However the findings do not necessarily reflect either the type of population or instruction; another prison population (Schrum, 1985) experienced a considerable number of positive benefits and CAI was reported as a popular method in another study (Nicol and Anderson, 2000).

Table 4: Wider benefits of learning in included RCTs

P = reported by participants   T = observed by teachers (?) = source unclear

Auothor(s), year	Intervention	Wider benefits
Batchelder and Rachal (2000 a and b)	Evaluation of the effect of CAI in reading and maths skills to male prisoners participating in a prison education programme.	Dislike of instructional method (without teacher support) and of testing (?)
Bean and Wilson (1989)	Investigation of the use of closed captioned television to develop sight vocabulary of adults attending literacy programme.	Enjoyment of instructional method (cf. traditional print instruction) (T) Self-reported gains in learning (P)
Martinson and Friedlander (1994)	Evaluation of educational effects of a GAIN programme on welfare recipients.	Mainly positive attitude of staff to instructional method (T) Possible negative attitude of participants to education/TALS testing (?) Increased use of job search services and work experience Increased receipt of GED Increased participation in education (P)
Nicol and Anderson (2000)	Comparison of computer-assisted vs teacher-directed instruction in numeracy to adults with mild learning disability.	Enjoyment of instructional method (CAI) (T) Transfer of skills to other contexts (?)
Rich and Shepherd (1993)	Investigation of benefit of teaching text comprehension strategies to improve comprehension of adult poor readers.	Instructional method favourably regarded by students (P/T)
Schrum (1985)	Evaluation of the effectiveness of an instructional strategy for accelerative and mastery learning of basic arithmetic by male prisoners referred for academic remediation.	Transfer of skills to other contexts. Increased self-esteem Improved life-skills (P?)
St. Pierre et al. (1993)	Variety of activities that included adult education.	Instructional method well regarded – supportive staff and provision for childcare Increased self-esteem and confidence in ability to cope with life Improved interaction with children Slight increase in earnings Progress towards achievement of personal goals Self-reported gains in learning (All P)

Transfer of skills learned to other contexts was noted in both of the numeracy studies and there were two instances of self-reported gains in learning (Bean and Wilson, 1989; St. Pierre et al. 1993). The only two mentions of improved self-esteem were as the result of an intervention that used holistic methods of relaxation and expectation of success (Schrum, 1985) and in a study that specifically considered the social implications of the intervention (St. Pierre et al. 1993). It was also claimed in Schrum that the benefits extended to participants' perceptions of an improvement in their life skills and change to a crime-free life. Additional benefits were also claimed in St. Pierre et al. to include increased ability to cope with the demands and stresses of life, more effective interaction with their children, progress towards personal goals and self-reported gains in learning. Thus the two studies which noted increased self-esteem and confidence also claimed the widest range of other positive benefits for the participants.

Issues of participants' confidence and self-esteem were also raised in Martinson and Friedlander (1994) but only to note their probable absence in the particular study population of welfare recipients. The view expressed that attendance in separate basic education classes

might isolate or stigmatise GAIN participants was never subsequently followed up.

The validity of the above findings must be viewed in the context of considerable methodological shortcomings. With the exception of St. Pierre et al. (1993), none of the studies stated that the collection of qualitative data was an objective; however, if such data are reported within a study, then rigour of methodology must apply as much to qualitative as to quantitative data.

The reporting of data on the wider benefits of learning was affected by the following weaknesses in particular. The source of the indicators was not adequately described or exemplified in any study (except St. Pierre et al. 1993); in particular it was not always clear whether reporting was by teachers or participants. For example, to state that 'interviews with participants and staff' took place (Batchelder and Rachal, 2000 a and b) but not subsequently to match the information to its source is unhelpful. It was never clear what form the 'attitude survey' used in Bean and Wilson took, and the other studies were even vaguer, for example, referring to feelings 'noted' by participants (Schrum, 1985).

With the exception of Martinson and Friedlander (1994), the proportion of participants to whom the indicators applied was not adequately described. Even when a study claimed 100 per cent enjoyment of the method of instruction (Bean and Wilson, 1989) it was simultaneously noted that the more able students (proportion never stated) were 'less positive'. Otherwise reference was made in vague terms to 'some' of the participants (Schrum, 1985), 'many' (Batchelder and Rachal, 2000 a and b), 'majority' (Bean and Wilson, 1989), or 'a number of' (St. Pierre, et al.1993); some findings were stated without any quantifier.

The St. Pierre et al. study at least gave some information about the questions asked in the interviews with parents and made it clear that all the benefits were those reported by the participants.

Our overall conclusion was that there was very little of value regarding the wider benefits of learning in the evidence from these RCT reports.

4. Discussion

4.1 Amount and quality of evidence
In this systematic review we found only nine relevant RCTs of interventions for adult literacy or numeracy. Given that low adult literacy and numeracy is an important problem (socially and individually) and that large amounts of public and private resources are devoted to addressing the issue, it is surprising that so few trials have been undertaken. Furthermore, it is worrying that the few trials that have been undertaken tend to be of low methodological quality, which renders their conclusions unreliable.

Within the RCT literature there is substantial heterogeneity between the included studies both in terms of the settings in which the studies took place and the interventions being evaluated. This heterogeneity makes it difficult to draw either quantitative or qualitative conclusions about which particular forms of intervention are effective. Furthermore, there appears to be substantial publication bias, with negative studies or studies with very low effect sizes not included, as evidenced from the funnel plot. Had these studies been available, the tentative

discovery of effects from adult literacy and numeracy programmes might not have held true.

Only one study, the California GAIN evaluation (Martinson and Friedlander, 1994), was large enough to detect small, but important, improvements in literacy and numeracy among the participants. Unfortunately, this study had severe methodological shortcomings, such as problems with random allocation. Further, this study also used non-standardised interventions among the participants, making it difficult to know whether individual components of the intervention were effective. This would also make replication of this study difficult.

4.2 Wider benefits

The importance of accommodating participants' perspectives in adult literacy and numeracy education is often recognised in the literature. For example, Brooks et al. (2001a) note 'the quest for a positive self-image' as a major factor influencing students' participation in basic skills provision. Problems of motivation are crucial in the drive to reduce the number of adults with low levels of literacy and numeracy. However, there is also the perception that issues of individual rights and wider benefits are in danger of being marginalised in what is seen as the competing discourse of increasing standardisation and testing of basic skills (Hamilton et al. 2001).

The fact that questions of confidence and self-esteem were raised in only a minority of the included studies is a surprising finding, in view of the importance placed on these by many basic skills educators. However, too much significance should not be attached to this finding; the inconsistent quality of the included trials has already been noted and there was no attempt at a rigorously systematic collection of qualitative data. At the same time, it is important to realise that any attempt to assess gains in confidence can be problematic and misleading. Mace (1992) warns against 'making claims for … confidence gain being a life change' and also notes the phenomenon of what she calls 'generosity' on the part of the student. 'Wishing to be generous to us, as tutors or literacy organisers, for what she saw as our generosity in taking her seriously, the literacy student gave us the compliment of attributing this change in herself to us – and to literacy itself' (Mace, 1992: 10). This 'generosity' problem may well apply to all participant self-reporting. It is possible that collection of such data by people other than the students' tutors would help to minimise this bias.

5. Conclusions

5.1 Impact on learning

There may be some evidence that adult literacy and numeracy programmes are more effective than no teaching. This finding is reassuring. Whilst this finding is based on three relatively small trials it is important that the most rigorous trial that we identified in the whole systematic review supported this finding (Rich and Shepherd, 1993). This same trial also provided evidence that merely exposing students to educational materials, without teaching, was not more effective than not exposing students to either teaching or the materials. This suggests, therefore, that 'self-study' may not be an appropriate strategy in this group of adults.

5.2 Poor quality of included trials

Because the trials are of relatively poor quality and very heterogeneous it is difficult to make

a conclusion about the effectiveness of a specific intervention. Similar conclusions are often reached in systematic reviews of research in other disciplines. For example, in health services research it is not unusual for reviews to be unable to make definitive statements about the effectiveness of an intervention because the included studies are of such poor quality. For example, a recent systematic review of randomised trials of over-the-counter cough medicines for acute coughs concluded as follows (Schroeder, 2002):

> *"Our results should therefore be interpreted with caution. Potential sources of bias such as randomisation procedure, blinding of outcome assessment, and losses to follow up were inadequately reported in several studies, suggesting poor methodological quality ....."*

5.3 Wider benefits

On the basis of the limited data available there is no evidence that the range of wider benefits of learning increases in proportion to the apparent effectiveness of an intervention. The two studies where the positive effect of the intervention was found to be statistically significant (Schrum, 1985; Rich and Shepherd, 1993) showed a large and a small number of wider benefits respectively. In the case of Schrum there is a need to assess the extent to which the wider benefits are attributable to the holistic method of instruction rather than to the actual progress made in numeracy. The study in which the greatest range of wider benefits was reported (St. Pierre et al. 1993) emphasised the participants' very positive perceptions of their teachers as supportive and respectful. There is also some evidence from these studies that increased self-esteem and confidence of participants is linked to an increased range of wider benefits.

6. Recommendation

It is recommended that a series of large well-designed, well-conducted and well-reported RCTs is funded in the field of adult literacy and/or numeracy. An outline proposal for one such trial is presented in Appendix I.

A systematic review of controlled trials (CTs)

1. Introduction

1.1 The research question
The research question for the systematic review of the CTs was: What is the evidence for the effectiveness of interventions or pedagogies for increasing literacy or numeracy in adults in studies using a quasi-experimental design?

1.2 Objectives
The primary objectives of this review were:

- to locate, synthesise and quality appraise all the CTs intended to evaluate interventions in adult literacy and/or numeracy and published between 1980 and 2002; and
- to establish what conclusions, if any, could be drawn about the effectiveness of teaching.

2. Methods

Systematic review methods were used throughout, in order to minimise bias. Studies were systematically searched for and located. They were then screened for inclusion in the mapping section of the review using pre-established inclusion and exclusion criteria. Finally, the included CTs were data extracted, quality appraised and synthesised. At the screening and data extraction stages of the review two researchers worked independently; they then conferred and agreed on all decisions.

The rest of the methodology of the project is given in full in Appendix C

3. Results

Nine controlled trials were independently double data extracted and quality appraised for this systematic review. The quality of the controlled trials was variable, but many of the studies had substantial methodological weaknesses. There was little evidence of publication bias.

3.1 Description of included CTs
Of the nine CTs included, six evaluated interventions in literacy and three evaluated interventions in literacy and numeracy:

Table 1: Study topics of included CTs

Literacy	Literacy and Numeracy
Dietrich, 1994	Diem and Fairweather, 1980
Gretes and Green, 1994	Lavery et al. 1998
Maclay and Askov, 1988	Meyer, 1983
Nurss, 1989	
Roberts et al. 1994	
Schrader, 1984	

3.2 Methodological quality of studies

All of the trials included were of a high quality in the sense that they adopted an appropriate design for questions of effectiveness, i.e. controlled trials, although this study design is always inferior to the RCT because it cannot control for all known and unknown variables that could influence outcome. Also, all of the studies included reported sufficient data for the effect sizes to be calculated by the reviewers. In a number of studies equivalence at baseline was established; and in some cases where the groups were not equivalent at baseline the pre-test scores were used as a covariate in the analysis in order to control for this problem. In nearly all cases discussion of study findings was consistent with the data (eight out of nine trials).

Table 2 shows whether or not the studies fulfilled our modified CONSORT checklist of quality criteria. As the table shows, the quality of the trials tended to be problematic, only one trial using intention to teach analysis (with this information being unclear in two other studies). It was not possible to ascertain whether or not follow-up measures were administered blind (unclear or not stated in all nine trials). Detailed information on the CTs is tabulated in Appendix F; and in Appendix G we consider the quality of each trial in depth.

3.3 Some methodological problems in the CTs

All of the following methodological problems occurred in at least one of the CTs; some problems occurred in more than one of the studies included.

(1) Groups were randomly assigned to intervention and control conditions; however too few groups were involved in the trials (two in two-armed trials or three in three-armed trials). Insufficient groups in a group or cluster-randomised trial means that any characteristic that occurs in one group cannot be counter-balanced with its presence in another group. A single group in each trial arm can never balance out any known or unknown co-variate; therefore, such a study is characterised as being a controlled trial.

(2) Attrition can be large or not stated. Any attrition can lead to bias, but this possibility is increased where the attrition is differential between the two groups, leading to the possibility of a biased outcome and an over- or under-estimate of the effectiveness of the intervention. Attrition bias can be minimised by continuing to gather outcome data, even though participants are no longer receiving the intervention and including all participants in the analysis in their original group allocation (intention-to-teach analysis).

(3) A lack of equivalence at baseline and this was not taken account of within the analysis.

(4) A lack of matching on pre-test scores (pre-test scores are an important covariate) and groups may differ on other unknown covariates, which may influence outcome. Whilst for a non-randomised controlled trial we can never be sure that selection bias has been eliminated, if the groups are balanced on the main outcome measure (e.g. numeracy scores) we can place more reliance on the results of a balanced study compared with an unbalanced trial or one that does not present baseline data.

(5) A lack of 'blinding' or 'masking' of the group allocation to the outcome assessor. This is important in order to minimise bias in the assessment that can lead to an over- or under-estimate of the effectiveness of the intervention.

(6) Not reporting on all outcome measures as stated in the methods section. Failure to report some of the outcome measures can lead to the suspicion of 'selective' reporting, that is only reporting the ones that show a positive effect.

Table 2: Quality of CTs included

Y = Yes N = No

N/S = not stated U = unclear

	Diem and Fairweather, 1980	Dietrich, 1994	Gretes and Green, 1994	Lavery et al. 1998	McClay and Askov, 1988	Meyer et al. 1983	Nurss, 1989	Roberts, 1994	Schrader, 1984
Was the study population adequately described? (i.e. were the important characteristics of the adults described e.g. age, gender?).	N	Y	Y	Y	Y	Y	Y	Y	N
Was the minimum important difference described? (i.e. was the smallest educationally important effect size described?).	N/S	N/S	N/S	N/S	N/S	N/S	N/S	N/S	N/S
Was the target sample size adequately determined?	N/S	N/S	N/S	N/S	N/S	N/S	N/S	N/S	N/S
Was intention to treat analysis used? (i.e. were all adults included in the follow-up and analysis?).	N	N	U	N	N	N	N	U	Y
Was the unit of allocation described (i.e. individual adults or groups of adults)?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Were follow-up measures administered blind? (i.e. were the researchers who administered the outcome measures blind to treatment allocation?).	N/S	N/S	N/S	U	N/S	N/S	N/S	N/S	N/S
Was estimated effect on primary and secondary outcome measures stated?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Was precision of effect size estimated (confidence intervals)?	N/S	N/S	N/S	N/S	N/S	N/S	N/S	N/S	N/S
Were summary data presented in sufficient detail to permit alternative analyses or replication?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Was the discussion of the study findings consistent with the data?	N	Y	Y	Y	Y	Y	Y	Y	Y

Table 3:   Description of CTs included

Country	Population	Intervention	Effectiveness
Diem and Fairweather, 1980, USA	Male inmates in adult detention center	Use of computer-assisted educational system for literacy and numeracy	No difference between intervention and control
Dietrich, 1994, USA	Adults registered for a reading and study skills course at a community college	The use of auditory perception training on reading ability	No difference between intervention and control for two outcomes; positive effect for control group on one outcome
Gretes and Green, 1994, USA	Student volunteers in adult basic education in community colleges	The use of computer-assisted instruction for reading	Large positive effect for intervention group
Lavery et al. 1998, New Zealand	Adults not in paid employment attending government-funded employment training programme	The use of computer-assisted instruction for basic literacy education	No difference between the two groups
McClay and Askov, 1988, USA	Parents of children receiving government-funded remedial programmes in reading and maths in low income areas	The use of computer-assisted instruction program for literacy	Positive effect for the intervention
Meyer et al. 1983, USA	Adult male inmates in maximum and medium security prisons	The use of computer-managed instruction for literacy and numeracy	No difference between the two groups
Nurss, 1989, USA	Male and female adults at an adult education centre	Use of computer-assisted instruction for reading	Positive effect for the control
Roberts et al. 1994, USA	Adult male prisoners in medium security prison	The use of a community building programme for reading	Positive effect for the intervention
Schrader, 1984, USA	Full-time technical institute students	The use of a developmental remedial reading programme	No difference between the two groups

Eight of the nine CTs included were undertaken in the USA, whilst the remaining study was undertaken in New Zealand. As Table 3 shows, three of the nine trials showed a positive effect for the interventions, five trials showed no difference and one trial showed a positive effect for the control treatment.

Table 3 also shows that the interventions and the study populations were very heterogeneous. This heterogeneity precludes a meta-analysis of all the studies, because interventions that may be effective in one setting, a non-prison population, for example, may not be as effective within another setting, for example a prison setting.

Figure 1: Funnel plot of controlled trials
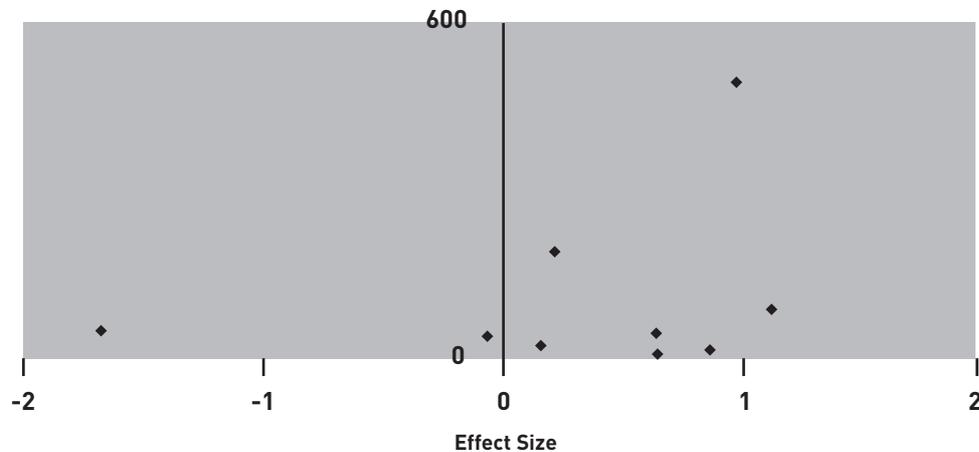
Figure 1 shows a funnel plot of the CTs . The figure is less suggestive of publication bias compared with the funnel plot of the RCTs.

4. Discussion

4.1 Amount and quality of evidence

In this systematic review we found only nine relevant CTs of interventions for adult literacy or numeracy where reviewers were able to calculate the effect sizes from the data included in the studies. It is of concern that the few trials that have been undertaken tend to be of low methodological quality, which renders their conclusions unreliable.

As with the randomised trials the literature is dominated by one country: the USA. Only one study, from New Zealand, was conducted outside the USA. Therefore, even if the trials had been of high rigour it would have been difficult to extrapolate their findings outside North America given the lack of replication in other English-speaking nations.

Within the CT literature there is substantial heterogeneity between the included studies both in terms of the settings in which the studies took place and the interventions being evaluated. This heterogeneity makes it difficult to draw either quantitative or qualitative conclusions about which particular forms of intervention are effective.

5. Conclusions

The CTs tended to be of poorer reporting quality than the RCTs. Given the limitations of the quasi-experimental design (i.e. it cannot control for all known and unknown extraneous variables) it is crucial that researchers adopting this study adopt methods to:

■ minimise attrition bias by gathering data on all original participants and including all participants in the analysis in their original group allocation;

■ use strategies to ensure equivalence at baseline or, if equivalence cannot be established, use pre-test scores as covariates in the analysis;

■ present baseline data for all groups;

■ 'blind' or 'mask' group allocation to the outcome assessor(s) and

■ report on all outcome measures as stated in the methods.

A number of the CTs used a control group that had significantly different pre-test scores. This choice of an inadequate control group meant that in some trials the differences between the post-test means of intervention and control groups were very similar and did not reflect changes in gain scores from pre-test.

There have been few attempts to expose common adult literacy or numeracy programmes to rigorous evaluation and therefore in terms of policy and practice it is difficult to make any recommendations as to the type of adult education that should be supported. In contrast, however, the review does provide a strong steer for educational researchers. Because of the present inadequate evidence base rigorously designed RCTs are required as a matter of urgency.

References

Included Studies * RCTs ** CTs

Adult Literacy and Basic Skills Unit (1994). **Basic skills research. Bibliography of research in adult literacy and basic skills 1972-1992.** London, Adult Literacy and Basic Skills Unit.

Argento, B.J. et al. (1980). **Alternative education models – preliminary findings of the Job Corps educational improvement effort. Education and training approaches.** Youth knowledge development report 5.2, Employment and Training Administration, District of Columbia, USA.

Askov, E.N. et al. (1986). **Evaluation of computer courseware for adult beginning reading instruction in a correctional setting. Final report. Pennsylvania, Pennsylvania State University.

*Batchelder, J.S. and J.R. Rachal (2000a). "Effects of a computer-assisted-instruction program in a prison setting: An experimental study." **Journal of Correctional Education** 51: 324-32.

*Batchelder, J.S. and J.R. Rachal (2000b). "Efficacy of a computer-assisted instruction program in a prison setting: An experimental study." **Adult Education Quarterly** 50: 120-33.

*Bean, R.M. and R.M. Wilson (1989). "Using closed captioned television to teach reading to adults. " **Reading Research and Instruction** 28: 27-37.

Beder, H. (1999). **The outcomes and impacts of adult literacy education in the United States** National Center for the Study of Adult Learning and Literacy, Massachusetts, USA.

*Boudett, K.P. and D. Friedlander (1997). "Does mandatory basic education improve achievement test scores of AFDC recipients?" **Evaluation Review** 2: 568-588.

Bouffard, J.A., Mackenzie, D.L. and L.J. Hickman (2000). "Effectiveness of vocational education and employment programs for adult offenders: A methodology-based analysis of the literature." **Journal of Offender Rehabilitation** 31: 1-42.

Broughton, G. (1994). **A test and control experiment that will assess the effectiveness of sensory-bridging, computer hardware and software, in aiding visually-impaired, deaf, and dyslexic students in achieving greater academic gains as reflected by attainment of proportionately higher scores on simulated versions of the standardized TABE test and a simulated GED examination. USA, Florida.

Broussard, R.L. (1983). **Home-based computer-assisted adult education project-phase III. Final project report Louisiana, University of Southwestern Louisiana.

Burtoff, M. (1985). **Haitian Creole literacy evaluation study – final report. Center for Applied Linguistics, USA, District of Columbia.

*Cheek, E.H. and J.D. Lindsey (1994). "The effects of two methods of reading instruction on urban adults' word identification and comprehension abilities." **Journal of Vocational and Technical Education** 11: 14-19.

Conti, G.J. and W.E. Giles (1980). **Employment through lifelong learning. A project demonstrating long-term simultaneous training in academic, vocational, and self-awareness skills –final report. Project operation dates: 1 January 1980 to 30 September 30 1980. Illinois: Spoon River College.

Cox, J.L. and C. Lane (1981). **Evaluation of the adaptation of the personalized system of instruction to non-traditional adult learners –final report. Volume I: Executive summary. North Carolina, Research Triangle Institute.

Culclasure D.F. (1982). **A pilot project to evaluate the use of low-cost microcomputers to improve the effectiveness of ABE services provided mentally handicapped adults. Texas: San Antonio State Hospital and State School.

Diem, R.A. and P.G. Fairweather (1980). ''An evaluation of a computer-assisted education system in an untraditional academic setting – A county jail.'' **AEDS Journal 13: 204-13.

Dietrich, J.A. (1994). **The effect of auditory perception training on the reading ability of adult poor readers. Rhode Island, USA

Dirk, J.M. and M. Crawford (1993). ''Teaching reading through teaching science: Development and evaluation of an experimental curriculum for correctional ABE programs.'' **Journal of Correctional Education 44: 172-76.

**Enriquez, B.E. et al. (1990). Family literacy for parents of pre-schoolers: A title VII first-year evaluation report. Texas, Education Service Center Region.

Gretes, J.A. and M. Green (1994). ''The effect of interactive CD-ROM/digitised audio courseware on reading among low-literate adults.'' **Computers in Schools 11: 27-42.

Irby, T.R. et al. (1992). **The Joliet Junior College Center for Adult Basic Education and Literacy's ''Families about Success'': Intergenerational programming that works. Illinois, Joliet Junior College.

Jancic, M. (1998). ''Does correctional education have an effect on recidivism?'' **Journal of Correctional Education** 49: 152-161.

Johnston, S. (1980). **The use of visual imagery to teach reading comprehension to adult education students. California, USA.

Kemple, J.J. et al. (1993). **The National JTPA Study: Site characteristics and participation patterns.** New York, Manpower Demonstration Research Corporation.

Kruidenier, J. (2002). **Research-based principles for adult basic education reading instruction.** Washington DC, National Institute for Literacy.

Kulik, C.-L. C et al. (1986). ''The effectiveness of computer-based adult education: A meta-analysis.'' **Journal of Educational Computing Research** 2: 235-52.

Lavery, L., Townsend, M. and K. Wilton (1998). ''Computer-assisted instruction in teaching literacy skills to adults not in paid employment'' **New Zealand Journal of Educational Studies 33, 2, 181-192.

Lawrence, D. (1988). **Enhancing Self-Esteem in the Classroom**. London: Paul Chapman

Maclay, C.M. and E.N. Askov (1988). ¨Computers and adult beginning readers: An intergenerational approach.¨ **Lifelong Learning 11: 23-28.

Macmurdo, A. (1988). **Technology Literacy Project. Final financial status and performance report, F.Y. 1988. Louisiana: Iberia Parish Library.

*Martinson, K. and D. Friedlander, (1994). **GAIN: Basic education in a welfare-to-work program. California's Greater Avenues for Independence Program.** New York: Manpower Demonstration Research Corporation.

McCrossan, L.V., McDowell, C. and P. Cooper (1998). **A comparison of the effectiveness of open entry/open exit classes to closed entry classes requiring attendance. Final report 1997–1998.** Pennsylvania: Adult Literacy Center of the Lehigh Valley.

*McKane, P.F. and B.A. Greene (1996). ¨The use of theory-based computer-assisted instruction in correctional centers to enhance the reading skills of reading-disadvantaged adults.¨ **Journal of Educational Computing Research** 15: 331-344.

Meyer, L.A. et al. (1983). **Evaluation research in basic skills with incarcerated adults. Technical report no. 303. Illinois: Illinois University

Morrow, D.H. et al. (1993). **Effectiveness of group instruction in adult literacy acquisition. Final report. Pennsylvania: Greater Pittsburgh Literacy Council.

Newsom, R. and L. Foxworth (1980). **A study of the impact of specific teaching strategies on locus of control of adult students, July 1979 to June 1980. Texas: North Texas State University.

*Nicol, M.M. and A. Anderson (2000). ¨Computer-assisted vs. teacher-directed teaching of numeracy in adults.¨ **Journal of Computer Assisted Learning** 16: 184-192.

Nurss, J.R. (1989). **PALS Evaluation Project. Georgia: Georgia State University.

Rachal, J.R. (1995). ¨Adult reading achievement comparing computer-assisted and traditional approaches: A comprehensive review of the experimental literature.¨ **Reading Research and Instruction** 34: 239-58.

Ramsey R.S. and R.W. Ramsey (1983). **The learning disabilities program in Job Corps. USA: Antlanta, Georgia.

*Riccio, J. et al. (1994). **GAIN: Benefits, costs, and three-year impacts of a welfare-to-work program. California's Greater Avenues for Independence Program.** New York: Manpower Demonstration Research Corporation.

*Rich, R. and M.J. Shepherd (1993). ¨Teaching text comprehension strategies to adult poor readers.¨ **Reading and Writing** 5: 387-402.

Roberts, R.E. and E.H. Cheek (1994). ¨Group intervention and reading performance in a medium-security prison facility.¨ **Journal of Offender Rehabilitation 20: 97-116.

Scanlon, D., Mellard, D.F., S. Garrison, et al. (1998). **What we know about literacy practices for adults with learning disabilities: A review of published research.** USA: Massachusetts

Schrader, M.A. (1984). **Effect of developmental/remedial reading program participation on the grade point average of students with reading skills deficiencies at Lakeshore Technical Institute. USA: Wisconsin.

*Schrum, H.E. (1985). **A modified instructional strategy for accelerative and mastery learning of basic arithmetic by male incarcerated adults.** USA: California.

Smith, M. and Z. Dalheim (1990). **Project READ. A study of twenty reading disabled adults. Massachusetts: Pittsfield Public Schools

St. Pierre, R.G., Swartz, J.P., Gamse, B., Murray, S., Deck, D. and P. Nickel (1993). **National evaluation of the Even Start Family Literacy Program: Final report.** Cambridge, Massachusetts: Abt Associates Inc.

Sticht, T.G. (1995). **The military experience and workplace literacy: A review and synthesis for policy and practice.** Pennsylvania: National Center on Adult Literacy.

Tewksbury, R.A. (1994). ''Improving the educational skills of jail inmates: preliminary program findings.'' **Federal Probation 58: 55-59.

Thuy, V.G. (1992). **High-tech for effective ESL/family literacy instruction. Final report. Pennsylvania: Indochinese-American Council.

USA. Lehigh County Community College. (1993). **Verifying applied literacy skills (VALS) in ABE programs. Pennsylvania.

USA. Indiana O.I.C. of America State Council (1990). **A comparative study of adult education Indianapolis/Richmond. Third-Party Evaluator. Final report. Indiana.

USA. Mid Continent Regional Educational Laboratory (1991). **Incorporating audio support into English composition CAI for adult learners. Phase II final report. New Mexico.

USA. Rio Salado Community College (1991). **SALSA (Southwest Advanced Learning System for Adults). Pilot project research report. Arizona.

**Wadsworth, S.G. and W.D. Frazier (1982). PLATO use in Oklahoma skill centers. Final report. Oklahoma: Oklahoma State Department of Vocational and Technical Education.

Wilson, A.M. (1992). **The INVEST Program: A computer-based system for adult academic upgrading. A pilot project. Canada: New Brunswick.

Wilson, D.B., Gallagher, C.A. and M.B. Coggleshall (1999). ''A quantitative review and description of corrections-based education, vocation, and work programs.'' **Corrections Management Quarterly** 3: 8-18.

Wisher, R.A. and J.W. O'Hara (1981). **Computer-based approach to the Navy's academic remedial training, Project Prest: A cost-effectiveness evaluation. San Diego, California: Navy Personnel Research and Development Center.

Other references

Albanese, M. (2000). ''Problem-based learning: why curricula are likely to show little effect on knowledge and clinical skills.'' **Medical Education** 34: 729-738.

Basic Skills Agency (1997). **Staying the Course: the Relationship Between Basic Skills Support, Drop Out, Retention and Achievement in Further Education Colleges.** London: Basic Skills Agency.

Basic Skills Agency (1998). **Family Numeracy Adds Up: Lessons from the Family Numeracy Pilot Programme.** London, Basic Skills Agency.

Basic Skills Agency (2000). **Effective Basic Skills Provision for Adults**. London: Basic Skills Agency.

Beder, H. (1999). **The outcomes and impacts of adult literacy education in the United States (NCSALL Report 6).** Cambridge, Massachusetts, Harvard Graduate School of Education, National Center for the Study of Adult Learning and Literacy.

Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., Pitkin, R., Rennie, D., Schulz, K.F., Simel, D. and D.F. Stroup (1996). ''Improving the quality of reporting of randomised controlled trials: The CONSORT statement.'' **Journal of the American Medical Association**, 276: 637-639.

Brooks, G. (2002). **What works for children with literacy difficulties? The effectiveness of intervention schemes**. London: DfES Research Report no. RR380

Brooks, G., Davies, R., Duckett, L., Hutchison, D., Kendall, S. and A. Wilkin, (2001b). **Progress in Adult Literacy: Do Learners Learn?** London, Basic Skills Agency.

Brooks, G., Giles, K., Harman, J., Kendall, S., Rees, F. and S. Whittaker (2001a). **Assembling the Fragments: a Review of Research on Adult Basic Skills.** London: Department for Education and Employment Research Report no.220.

Brooks, G., Gorman, T., Harman, J., Hutchison, D., Kinder, K., Moor, H. and A. Wilkin (1997). **Family Literacy Lasts: the NFER Follow-up Study of the Basic Skills Agency's Demonstration Programmes.** London, Basic Skills Agency.

Brooks, G., Gorman, T., Harman, J., Hutchison, D. and A. Wilkin, (1996). **Family Literacy Works: the NFER Evaluation of the Basic Skills Agency's Demonstration Programmes.** London: Basic Skills Agency.

Brooks, G., Harman, J., Hutchison, D., Kendall, S. and A. Wilkin (1999). **Family Literacy for New Groups: the NFER Evaluation of Family Literacy with Linguistic Minorities, Year 4 and Year 7.** London: Basic Skills Agency.

Brooks, G., Heath, K. and A.C. Pollard (2004). **A review of assessment instruments for adult literacy and numeracy.** A report of the DfES National Research and Development Centre for Adult Literacy and Numeracy. London, National Research and Development Centre for Adult Literacy and Numeracy.

Coben, D. (2003). **Adult numeracy: review of research and related literature. A report of the DfES National Research and Development Centre for Adult Literacy and Numeracy.** Nottingham, University of Nottingham.

Comings, J. (2003). 'Evidence-based education: a perspective from the United States.' Paper presented at NRDC/CCER international conference, Nottingham, March.

Constable, H. and R. Coe (2000). ''Evidence and Indicators: Dialogue, Improvement and Researching for Others.'' **Evaluation of Research in Education** 14: 115-123.

Cook, T.D. and T.D. Campbell, (1979). **Quasi-Experimentation: Design and Analysis Issues for Field Settings.** Boston: Houghton Mifflin.

Cook, T.D. (2002). 'Reappraising the Arguments Against Randomised Experiments in Education: An Analysis of the Culture of Evaluation in American Schools of Education**.'** Paper for Presentation at the SRI International Design Conference on Design Issues in Evaluating Educational Technologies.

GB. DfEE (1999). **A fresh start: Improving adult literacy and numeracy (The Moser Report).** London: Department for Education and Employment.

GB. DfEE (2001). *Skills for life***.** London: Department for Education and Employment.

Hamilton, M., Macrae, C. and L. Tett, (2001). **Powerful literacies: the policy context.** In Crowther, J., Hamilton, M. and L. Tett, (Eds) Powerful Literacies. Leicester, NIACE.

Lipsey, M.W. and D.B. Wilson, (1993). ''The Efficacy of Psychological, Educational and Behavioral Treatment Confirmation From Meta-Analysis.'' **American Psychologist** 48: 1181-1209.

Mace, J. (1992). **Talking about literacy. Principles and practice of adult literacy education.** London, Routledge.

Mellar, H., Kambouri, M., Wolf, A., Goodwin, T., Hayton, A., Koulouris, P. and V. Windsor, (2001). **Research into the effectiveness of learning through ICT for people with basic skills needs**. Research report submitted to Ufi, June 2001.

Palincsar, A.S. (1982)**. Improving the reading comprehension of junior high school students through the reciprocal teaching of comprehension-monitoring strategies.** Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

Palincsar, A.S. (1986). ''The role of dialogue in providing scaffolded instruction.'' **Educational Psychology** 21: 73-98.

Palincsar, A.S. and A.L. Brown (1984). ''Reciprocal Teaching of comprehension-fostering and comprehension-monitoring activities.'' **Cognition and Instruction** 1(2): 117-72.

Petticrew, M. (2001). ''Systematic reviews from astronomy to zoology: myths and misconceptions.'' **British Medical Journal** 322: 98-101.

Rosenshine, B. and C. Meister (1994). ''Fostering literacy learning in supportive contexts.'' **Review of Educational Research** 64(4): 479-530.

Senn, S.J. (1989). ''Covariate imbalance and random allocation in clinical trials.'' **Statistics in Medicine** 8: 467-475.

Schroeder, K. and T. Fahey (2002). ''Systematic review of randomised controlled trials of over the counter cough medicines for acute coughs in adults.'' **British Medical Journal** 324: 329 - 31.

Sticht, T.G. and W.B. Armstrong, (1994). **Adult literacy in the United States: A compendium of qualitative data and interpretive comments.** San Diego, California: Applied Behavioral and Cognitive Sciences Inc and San Diego Community College District.

Torgerson C.J. and D.J. Torgerson (2003). ''The design and conduct of randomised controlled trials in education: Lessons from healthcare.'' **Oxford Review of Education** 29(1): 68-80.

Glossary

Attrition Some participants are often lost during the study and cannot be included in the analysis. This is called attrition.

Bias A term denoting that a known or unknown variable is or may be responsible for an observed effect, rather than the intervention.

Before and after study The least robust quasi-experimental design. Participants are given a pre-test, exposed to the intervention and then given a post-test with any differences being ascribed to the treatment.

Blocked randomisation This method of randomisation prevents groups becoming either numerically unbalanced or suffering from chance bias. It does this by randomising in blocks (for example a block of 4). Thus, a block of 4 can be: ABAB, AABB, BBAA, BABA. This means that the study will be balanced although the block size must be kept secret to conceal the allocation sequence.

CAI Computer-assisted instruction – where teaching and learning is supported by the use of computers.

Case control study A study with a control group that has been identified as being similar to the treatment group but was not formed by randomisation.

Concealed allocation This prevents the researcher, participant or teacher knowing in advance the allocation of an individual. This is important because random allocation can be undermined by choosing participants to be in a desired group, which subverts the trial.

CT Controlled trial – This usually means a study with a control group that has been formed by means other than randomisation; consequently the validity of the study is threatened by selection bias.

Co-variates or confounders These are variables that are associated with outcome. Randomisation is the only method that ensures both known and unknown co-variates are equally distributed among treatment groups.

CONSORT Consolidated Standards for Reporting Trials is the methodological standard adopted by many medical journals for publication of RCTs. In this report we have based our quality assessment on the CONSORT criteria.

Effect size When an outcome variable is measured on a continuous scale (e.g. changes in a test score) the improvement or decrement is described in standard deviation units, which is termed the effect size (see Table 1).

Table 1: Translating effect sizes into percentage improvement (Source: Albanese, 2000)

	Effect Size Improvement				
Percentile of Control	**0.20**	**0.30**	**0.50**	**0.80**	**1.00**
5th	7	9	13	20	26
10th	4	16	22	32	39
15th	20	23	30	41	48
25th	32	35	43	54	62
50th	58	62	69	79	84

The table above allows a translation of the effect size of an intervention into the average improvement for a class. For example, consider a test where the average mark for a class is 50 per cent. Let us suppose we have identified an intervention that has an effect size of 0.50 (or half a standard deviation). If we take the first column of the table and go to the 50th percentile and then look across to where the bottom row intersects with the 0.50 effect size we can see that this will take the average mark up to 69 per cent (i.e. an improvement of 19 percentage points).

Funnel plot A method of assessing whether there is any publication bias. The effect size of each study is plotted against its sample size. Small studies will have large random variations in their effect sizes, which will be scattered along the x-axis close to the bottom of the y-axis. Larger studies will be higher up on the y-axis and less scattered along the x-axis. A review with no publication bias will show a plot in the shape of an inverted funnel.

Intention-to-teach (ITT) analysis This is where all participants are analysed in their original randomised groups; it is the most robust analytical method.

Multi-variate analysis In an RCT most known and unknown variables affecting outcome will be balanced at baseline. Nevertheless, particularly in small studies, variables can still affect the precision of the results. This is particularly the case with the pre-test variable, which will strongly predict outcome. A more precise estimate of the effect size (i.e. with smaller confidence intervals) can be obtained by undertaking a multivariate analysis with the pre-test score as a co-variate as well as the group allocation.

Numbers needed to teach (NNT) This is a method of converting the effects of an intervention into an easily understood metric. Thus, an NNT of 5 means that 5 students need to be given an intervention for 1 to pass a set test or exam. See Table 2.

Table 2: Effect size and numbers needed to teach (NNT)

Mean Effect Size	Number of Extra Students passing a 50% test threshold	Numbers Needed To Teach
0	0%	–
0.1	4%	25
0.2	8%	13
0.3	12%	8
0.4	16%	6
0.5	19%	5
0.6	23%	4*
0.7	26%	4
0.8	29%	3*
0.9	32%	3
1.0	34%	3*

*Rounded to the nearest whole number.

Table 2 converts effect sizes into NNT (Torgerson and Torgerson, 2003). If for example an intervention has an effect size of 0.1 and an important exam pass threshold is 50 per cent, then by using the intervention 4 per cent extra students would pass the exam. In other words 25 students would need to be exposed to the new intervention to enable one more student to pass the exam.

Paired randomisation This is a commonly used method in educational research. Participants are formed into matched pairs on the basis of important co-variates (e.g. gender and/or pre-test scores). Once the study group has been formed into pairs a random member of each pair is allocated to the intervention.

Publication bias Not all RCTs are published. There is a well-established tendency for trials that produce negative effects or no effects to be less likely to be published than positive trials. Unless a systematic review includes these negative trials it can give a misleading optimistic assessment of the intervention. Existence of publication bias can be detected by using funnel plots.

RCT – Randomised Controlled Trial This is where two or more groups have been formed through random allocation (or a similar method). This is the only method that ensures that selection bias is eliminated at baseline.

Regression to the mean This statistical phenomenon occurs when test results are, by chance, some distance away from the mean. Consequently at post-testing the 'extreme' results will tend to regress to the mean. When selecting participants on extreme test results (e.g. very poor pre-tests) there will be an apparent dramatic improvement on post-test because of this effect (irrespective of the teaching method). Randomisation automatically controls for regression to the mean effects. Nevertheless, it can still have an influence if the groups are unbalanced at baseline on pre-test scores. This imbalance can be adjusted for by a multivariate analysis.

Selection bias This occurs when groups are formed by a process other than randomisation and means that important factors that are associated with outcome differ between the groups before they are exposed to the intervention.

Simple randomisation This is the easiest form of randomisation, akin to tossing a coin. A disadvantage with simple randomisation is that with small studies ($\leftarrow$150) there is a high chance of having large numerical imbalance between the groups. Restricted forms of randomisation are often used to deal with this.

Appendix A

Key methodological features of well-designed randomised CTs

A.1 Concealed allocation

Fair randomisation will, on average, produce equivalent groups. Methodological studies undertaken primarily in health services research have shown that random allocation has sometimes been compromised. Studies examining the effect sizes of trials that have randomly allocated participants using a method that is difficult or impossible to compromise have tended to show different, usually lower, effect sizes than trials using allocation methods that are open to being compromised. Using 'open' randomisation methods such as using random number tables means that the investigator knows the next allocation in advance of it happening. Therefore, the next participant can be 'excluded' from the study if he or she does not possess certain 'desirable' characteristics. This can then lead to bias, which undermines the whole basis of random allocation. It is important, therefore, that the 'mechanics' of randomisation are clearly described to see if the study is susceptible to 'subversion' bias or not.

A.2 Simple versus stratified allocation

Simple randomisation is the least complex method of random allocation. It will, however, often lead to groups that suffer a numerical imbalance. Thus, simple randomisation of 30 participants, will often lead, by chance, to quite disparate group sizes (e.g. 10 versus 20 rather than 15 versus 15). Also there may be an imbalance in terms of important co-variates. To prevent these problems one can use stratified allocation, which leads to numerical equivalence and balance between known co-variates. Often studies that seem to have used simple randomisation with small sample sizes appear to be well balanced in numerical terms. Whilst exact numerical balance can be achieved by chance, in small studies this is unlikely. It may be that investigators when seeing an imbalance occurring, by chance, may 'top-up' the smaller group with the last few participants. If this has occurred then the basis of random allocation has been undermined, allowing bias.

A.3 Paired allocation

It is a common practice in the field of educational trials to undertake 'paired' randomisation. In a two-group trial this means participants are matched on the basis of one or two factors. For example, participants may be paired on the basis of age or gender and pre-test educational score. One member of each pair is then allocated to the test treatment. The consequence of pairing is that there should be exactly equal numbers in each group and the group should be exactly balanced in terms of the characteristics on which the pairing took place. The disadvantage of using pairs (or triplets in a 3-group trial) is that unless you have an even number of participants you will have to exclude the odd person, thus losing power. Also, if the co-variate you pair on (e.g. age) has an unusual relationship with outcome this cannot be explored in the analysis as the pairing eliminates all variation due to that co-variate.

A.4 Intention to teach analysis

Once participants have been allocated to their respective groups it is important that they remain in those groups for analysis, to avoid bias. A common, but incorrect, method is to exclude some participants after randomisation for a variety of reasons. One approach is to do what is termed 'an on-treatment analysis' – this is where only those participants who demonstrate treatment fidelity are included in the analysis. Unfortunately, this can lead to bias as those participants who complete treatment are likely to be different from those who do not. Intervention received analysis can therefore produce a biased result.

A.5 Sample size and power calculations

Trials in educational research commonly exhibit a Type II error. This is where the sample size is insufficient to show, as statistically significant, a difference that is educationally important. Reviews of educational interventions have shown that most interventions will, at best, only lead to an improvement in the region of half a standard deviation and quite often somewhat less. Statistical theory shows that to reliably detect (with 80 per cent power) half a standard deviation difference as statistically significant ($p = 0.05$) for a normally distributed variable requires a minimum sample size of 126 participants. Studies that are smaller than this run the risk of erroneously concluding that there was not a significant difference when actually there was. Therefore, a good quality study ought to describe the reasoning behind the choice of sample size.

A.6 Blinded follow-up

Ideally, to avoid ascertainment bias, follow-up post-tests should be given and marked blindly so that the researcher is unaware of the membership of the group. This prevents 'marking' bias where a researcher may consciously or unconsciously preferentially give better marks to the experimental group.

A.7 Use of confidence intervals

The point estimate of effect of any intervention will always be imprecise. The level of the imprecision is dependent upon the sample size and event rate in the treatment groups. The use of confidence intervals (usually 95 per cent, but sometimes 99 per cent or 90 per cent) reflects this imprecision in the study results. Thus, for example, a treatment that has an effect size of 0.50, but a confidence interval of −0.1 to 1.2 is not statistically significant but will indicate to the reader that there is a relatively high possibility that there is a beneficial effect of treatment in excess of 1 standard deviation. In this instance, one might consider doing a further, larger randomised trial. In contrast, if the point effect was 0.05 and the confidence interval was −0.1 to 0.12 then the reader might consider that it is unlikely even with a bigger trial that this intervention would show an effect that was educationally worthwhile (assuming the conduct of the trial in question was of high quality).

Appendix B

Protocol for the systematic review of RCTs

Project B1.1 of the National Research and Development Centre for Adult Literacy and Numeracy

Research question

What evidence is there of the effectiveness of interventions or pedagogies for increasing literacy or numeracy in adults?

Background

It appears that no search of databases for empirical investigations of adult literacy and numeracy investigations has yet been carried out, whether for randomised controlled trials (RCTs) or for experiments with less rigorous designs. In their review of the research literature in the field, Brooks et al. (Assembling the Fragments, DfEE, 2001) found no RCTs, but a preliminary scoping exercise for this project (on ERIC and PsycINFO on 27/12/'01 and 29/12/'01) has revealed a number of investigations that appear to qualify.

Methods

This will be a systematic review of the research evaluating the effectiveness of interventions or pedagogies in adult literacy and numeracy. At each stage of the review, methods will be informed by the NHS Centre for Reviews and Dissemination guidelines for undertaking systematic reviews (http://york.ac.uk/inst/crd/report4.htm). In addition, the Institute of Education's Evidence-Informed Policy and Practice Co-ordinating Centre (EPPI) guidelines and tools for data extraction and quality appraisal of randomised controlled trials in educational research will inform this review (http://eppi.ioe.ac.uk/ED.Handbook/HBTitle.htm). Double data extraction and quality appraisal will be undertaken and where there is disagreement between reviewers this will be resolved through discussion.

An expert panel will advise at each stage of the review process. This will comprise experts on adult education and systematic reviewing techniques.

Preliminary searches

The systematic review will focus on educational outcomes; therefore the main databases for this review will be the educational databases. These are: PsycINFO, ERIC (Educational Resources Information Center), C2SPECTR (The Campbell Collaboration's Social, Psychological, Educational and Criminological Trials Register) and SSCI (Social Science Citation Index). In addition, a database of the grey literature will be searched: SIGLE (System for Information on Grey Literature in Europe). Because there may be relevant trials undertaken in prisons the Criminal Justice Abstracts database will also be searched.

Preliminary searches (December 2001) on the main two educational databases revealed the likelihood of a small amount of literature. Searching on adult literacy and numeracy on PsycINFO led to 299 hits of which 25 were potentially relevant for the review. Searching on ERIC on adult literacy only led to 66 hits of which 16 were potentially relevant for the review.

Full searches

Full electronic searches will be undertaken on: PsycINFO; ERIC; C2SPECTR; SSCI; SIGLE and the Criminal Justice databases. These searches will be undertaken by an independent information consultant using key words developed by the review co-ordinator (Professor Greg Brooks) and one of the Research Fellows (Carole Torgerson). The full electronic searches will be undertaken for the years 1980 to the present.

Inclusion criteria

Papers will be included according to the following criteria established a priori:

- Type of intervention: any intervention aimed at increasing the literacy or numeracy of adults (18 years and over).
- Study design: randomised controlled trials; controlled trials.
- Outcomes: educational outcomes of literacy or numeracy.
- [1] Studies undertaken in English speaking countries and written in the English language.
- Papers (published or unpublished): 1980 to January 2002

Exclusion criteria

Papers will be excluded according to the following criteria established a priori:

- Type of Intervention: any intervention aimed at increasing the literacy or numeracy of adult (ESOL) students studying in universities in English-speaking countries.
- Study design: pre- and post-test, post-test and all non-intervention studies.

Procedure

On completion of the searches the citations identified will be imported into the reference manager Endnotes, which allows for easy de-duplication of records. The titles (and where available) abstracts will be scanned for relevance. This screening will be undertaken by two researchers working independently. The results will be compared. All potentially relevant studies will be retrieved.

The full papers will be assessed again for relevance, and all studies meeting the inclusion criteria will be tabulated. Studies meeting the criteria for data extraction will be extracted, using pre-specified data extraction sheets (to ensure that the information is recorded in a systematic way) and coded into tables. The tables will include information about: the aims of the study, the type of intervention or pedagogy being evaluated, the outcomes measured and the study methodology (e.g. type of RCT, study size). We will list (a) quality criteria (allocation concealment, baseline equivalence, follow up etc.) and (b) study information (setting, country, intervention, control, outcomes etc.).

Two further focuses for the search will be:

- studies that include a follow-up assessment of learners at some point after an intervention, in order to judge the extent to which gains while in provision are sustained afterwards (e.g. test–re-test, follow-up assessment or interim results)

---

[1] I.e. countries where English is spoken as a first language by a significant segment of the population - includes UK, Ireland, USA, Canada, Australia, NZ, Jamaica and other countries in the Caribbean, Gibraltar and S. Africa, exclude India, Pakistan, HK, Singapore, Bangladesh, Malaysia and China.

■ wider benefits of learning, for example self-reported gains in literacy and numeracy, enhanced self-confidence, enrolment on further courses of study, gaining employment and (for parents in family learning) greater involvement with their children's schools.

Both of these analyses will be conducted on the references that are turned up by the search for trials.

Outputs
■ A map of the research literature including tables of randomised controlled trials, controlled trials: interventions, outcomes and findings.
■ Full data extraction and quality appraisal of all included randomised controlled trials.
■ Critical analysis of wider benefits.
■ List of all included and excluded trials.
■ A report containing results summary, conclusions and recommendations.
■ Dissemination for practitioners.
■ A journal paper for peer review.

Timetable
February 2002: Full electronic searches; screening of titles and abstracts for possible inclusion; sending for papers.
March 2002: Sending for papers; screening of full papers for inclusion; mapping and start of data extraction and quality appraisal.
April 2002: Data extraction and quality appraisal of RCTs;
May 2002: Data extraction and quality appraisal of RCTs; draft final report.
June 2002: Paper for peer-reviewed journal; final report.

Review team members

Review Co-ordinator: Professor Greg Brooks, University of Sheffield.

Greg Brooks, M.A., PGCE, Ph.D.
Professorial Research Fellow
Professor Greg Brooks joined the University of Sheffield School of Education on 1 January 2001. He taught for 10 years in secondary and higher education, variously in France, Kenya, Essex and Northern Ireland. While doing an M.A. in Applied Linguistics at the University of Essex he developed a research interest in the question of phonological coding in silent reading, and took that as the subject of his Ph.D. at the University of Leeds.

For the 20 calendar years 1981-2000 he worked at the National Foundation for Educational Research in Slough. In 1981-89 he worked full time on the Assessment of Performance Unit Language Monitoring Project – his main responsibility being the assessment of oracy (speaking and listening) skills of 11- and 15-year-olds. From 1990 to 2000 he directed over 20 research projects in the areas of language and literacy, including four surveys of schoolchildren's reading attainment, several research reviews, and seven evaluations of family and adult learning initiatives for the Basic Skills Agency.

He is a long-standing member of the UK Reading Association, including serving on the National Executive Committee (1991-2001) and on the Editorial Board of the Journal of Research in Reading since 1980. He was President of UKRA in 1999-2000. In 1991-98 he

represented UKRA on the International Development in Europe Committee (IDEC) of the International Reading Association, and he has been Chairperson of IDEC since 1998.

His main research interests are in language and literacy. He is currently directing the evaluations of the Peers Early Education Partnership (PEEP) in Oxfordshire, and of the Basic Skills Agency's Keeping Up with the Children project, which provides 12-hour courses for parents in England to help them understand the literacy and numeracy hours.

Publications

Brooks, G., Giles, K., Harman, J., Kendall, S., Rees, F. and S. Whittaker (2001). **Assembling the Fragments: a Review of Research on Adult Basic Skills**. London: Department for Education and Employment.

Brooks, G., Davies, R., Duckett, L., Hutchison, D., Kendall, S. and A. Wilkin (2001). **Progress in Adult Literacy: Do Learners Learn?** London: Basic Skills Agency.

Mifsud, C., Milton, J., Brooks, G. and D. Hutchison (2000). **Literacy in Malta: The 1999 National Survey of the Attainment of Year 2 Pupils**. Slough: NFER for the University of Malta.

Brooks, G., Harman, J., Hutchison, D., Kendall, S. and A. Wilkin (1999). **Family Literacy for New Groups: The National Foundation for Educational Research Evaluation of the Basic Skills Agency's Programmes for Linguistic Minorities, Year 4 and Year 7**. London: Basic Skills Agency.

Brooks, G., Flanagan, N., Henkhuzens, Z. and D. Hutchison(1998). **What Works for Slow Readers? The Effectiveness of Early Intervention Schemes**. Slough: NFER.

Contact: Prof. Greg Brooks/g.brooks@sheffield.ac.uk

Research Fellow: Carole Torgerson, University of York.

Carole Torgerson joined the University of York Department of Educational Studies in March 2001. She has a background in English teaching and research. She has completed a systematic review and meta-analysis on the effectiveness of ICT in the teaching and learning of spelling. She is currently working on a systematic review of the impact of ICT on literacy learning (directed by Professor Andrews). She is also collaborating on a project with the NHS Centre for Reviews and Dissemination (Dr Amanda Sowden) and the Centre for Research in Criminal Justice on a systematic review of the role of volunteers in education, health and the criminal justice system (funded by the Home Office).

Qualifications

Oct 2000:   Registration for Ed D. University of Sheffield, Department of Educational Studies.
1990:   Adv. Diploma in Special Educational Needs. Bishop Grosseteste College, Lincoln.
1983:   M.Litt. University of Stirling.
1980:   Post Graduate Certificate in Education. University of Nottingham.
1979:   B.A. (Hons) English and Related Literature 2:ii. University of York.

Publications

Torgerson, C. J. (2000). "The Erroneous Use of Statistics to Monitor Performance of Teachers and Schools. " **Curriculum** 21: 94-97.

Torgerson, C. J. (2001) "The Evidence for Misleading Statistics in Local Authority League Tables. " **Curriculum** 22: 26-29.

Torgerson, C. J. and D. J. Torgerson (2001) "The Need for Randomised Controlled Trials in Educational Research. " **British Journal Educational Studies** 49: 316-328.

Torgerson, C. J. and D.A. Elbourne (2002) "Systematic Review and Meta-Analysis of the Effectiveness of Information and Communication Technology (ICT) on the Teaching of Spelling." **Journal of Research in Reading** In Press.

Torgerson, C. J., and D. J. Torgerson (2003) "The Design and Conduct of Randomised Controlled Trials in Education: Lessons from Health Care." **Oxford Review of Education 2003**. Accepted for publication.

Torgerson, D. J., Torgerson, C. J., Birks, Y. F., J. Porthouse (2002) 'A comparison between the quality of randomised trials conducted in health care and education.' Unpublished conferences paper presented at Oxford Conference: Advances in Systematic Reviews. June 2002.

Contact: Carole Torgerson / cjt3@york.ac.uk

Research Fellow: Jill Porthouse, University of York.

Jill Porthouse joined the University of York Department of Health Sciences in 2001. She is trial coordinator for two large, multi-centred, randomised trials. These are a trial of hip protector plants in primary care to prevent fracture (funded by the National Lotteries Charity Board and the National Osteoporosis Society) and a trial of calcium and vitamin D supplementation to prevent fracture (funded by Northern and Yorkshire NHS Research and Development).

Qualification:

2002	BA (Hons) English Literature and History 2.1. St Martin's College, Lancaster.
1993	BSc Podiatric Medicine. Durham School of Podiatric Medicine.
1991	Diploma in Podiatric Medicine. Durham School of Podiatric Medicine.

Publications

Torgerson, D. J., Torgerson, C. J., Birks, Y. F., and J. Porthouse (2002) 'A comparison between the quality of randomised trials conducted in health care and education.' Unpublished conferences paper presented at Oxford Conference: Advances in Systematic Reviews. June 2002.

Jill Porthouse / jp38@york.ac.uk

Research Fellow: Maxine Burton, Manchester University

Maxine Burton has been involved in adult literacy since the early 1970s, working first as a volunteer tutor with Cambridge House in London and from 1985 to 2001 as a part-time basic skills tutor in further education (literacy and numeracy). In 2000/2001 she also delivered the literacy strand of the Keeping Up with the Children project at local primary schools.

She has taught at the University of Manchester since 1996, and is currently contributing to the

teaching, in the School of Education, of modules on literacy, child language acquisition and the structure of English (Grammar and Phonetics/phonology).

Her long-standing interest in language dates from her first degree in modern languages at Edinburgh University and for her M. Ling (General Linguistics) at Manchester University, she chose to specialise in child phonology. In 1998 she started an Ed. D. at the University of Sheffield, for which she is doing research into literacy in 19th century England, as a way of exploring the origins of the present-day stigma attached to illiteracy.

Publications

Burton, M. (2002). "The Stigma of Illiteracy." **Research and Practice in Adult Literacy Bulletin** 47.

Contact: Maxine Burton / maxineburton@btinternet.com

Research Secretary: Alison Robinson, University of York.

Alison Robinson is an experienced secretary/administrator who has worked in variety of environments in the private, voluntary and public sectors. Areas of expertise include developing and maintaining administrative systems and managing computerised applications.

She joined the University of York Department of Educational Studies in April 2001 as Secretary to the EPPI (Evidence-Informed Policy and Practice Initiative) English Review Group, which is conducting a systematic review of the impact of ICT on Literacy Learning (funded by the DfES). The focus of her role is to manage the bibliographic database and provide general administrative support for the Review Team. She has also assisted on a systematic review on the role of volunteers in education, health and the criminal justice system (funded by the Home Office) in collaboration with the University's NHS Centre for Reviews and Dissemination and the Centre for Research in Criminal Justice.

Contact: Alison Robinson / ar31@york.ac.uk

Independent Information Consultant: Kath Wright, York.

Kath Wright has been an information officer at the Centre for Health Economics/NHS Centre for Reviews and Dissemination at the University of York since 1997. She supports research staff in their work and, in particular, carries out literature searching for systematic reviews undertaken by staff within NHS CRD. She is also involved in the production and maintenance of DARE, a database of abstracts of systematic reviews that is available both on the web and on the Cochrane Library, and in providing training in literature searching at all levels.

Before working at York, she had worked in a number of special libraries including Merseyside Probation Service, the British Institute of Management and Manchester Business School. She has a Postgraduate Diploma in Librarianship and an MA in Library and Information Management Studies.

Review Consultant: Professor Ian Watt, University of York.

Posts held

1985–1986   Medical Officer (Primary Health Care), Andhra Pradesh, India, Action Health

1986–1987	Senior House Officer in Psychiatry, St Mary's Hospital, Scarborough, N Yorkshire
1987–1989	Principal in General Practice, Earby Surgery, Earby
1989	Medical Officer-Britain Nepal Medical Trust, Phidim/Illam, Nepal
1989–1993	Registrar/Senior Registrar in Public Health Medicine, Yorkshire Regional Health Authority (base district East Riding Health Authority)
1991–1993	Visiting Lecturer, Academic Unit of Public Health Medicine, University of Leeds
1994 –1999	Dissemination Manager, NHS Centre for Reviews and Dissemination, University of York
1994	Honorary Consultant in Public Health Medicine, North Yorkshire Health Authority
1995 – present	Part time GP Principal, Harewood Family Practice, Catterick Garrison, N Yorks
1997 – present	Professor of Primary and Community Care, Department of Health Studies, University of York

Qualifications

1978	BSc (Med Sci), St Andrews University
1981	MB ChB, Manchester University
1991	MPH, Leeds University
1993	Membership of the Faculty of Public Health Medicine
1999	Fellowship of the Faculty of Public Health Medicine

Publications

Parry, J. M., Foy, R. C., Woodman, C. B. J., Deeks, J. J., Watt, I. S. and T. A. Sheldon (1999) "A randomised controlled trial to assess the impact of a printed summary of research findings in general practice." **British Journal of General Practice** 49: 634-638

Entwistle, V. A., Watt, I. S. and F. Johnson (2000) "The case of Norplant as an example of media coverage over the life of a new health technology." **The Lancet** 355: 1633-36.

Scott, J. T., Entwistle, V. E., Sowden, A. J. and I. S. Watt (2001) "Giving tape recordings or written summaries of consultations to people with cancer: a systematic review." **Health Expectations** 4(3): 162-169.

Wilson, P., Droogan, J., Glanville, J., Watt, I. and G. Hardman (2001) "Access to the evidence base from general practice: a survey of general practice staff in Northern and Yorkshire Region." **Quality in Health Care** 10: 83-89.

Burr, R., Johanson, R., Wyatt, J., Watt, I. and P. Jones (2001) "A randomised trial of an intervention package designed to promote external cephalic version at term." **European Journal of Obstetrics and Gynaecology and Reproductive Biology** 100(1): 36-40

Expert panel:
Dr Andrea Nelson, University of York (systematic reviews)
Dr Lewis Pike, York (adult numeracy)

Appendix C

Full description of methodology for scoping and systematic reviews

1   Identification of studies

(a) Electronic searches
In collaboration with the two members of the review team (Greg Brooks and Carole Torgerson), an independent information consultant (Kath Wright) wrote the search strategies for the electronic databases and carried out electronic searches for the review on PsycINFO, ERIC (Educational Resources Information Center), SSCI (Social Science Citation Index), The Cochrane Library, C2-SPECTR (The Campbell Collaboration's Social, Psychological, Educational and Criminological Trials Register), SIGLE (System for Information on Grey Literature in Europe) and Dissertation Abstracts. Search terms used included adult literacy; adult numeracy; adult basic education; worksite education; workplace education (see Appendix D for search log and full strategies).

(b) Hand searches
All the bibliographies of included reviews were hand searched (CJT) for potentially relevant trials (RCTs and CTs) not already identified through the electronic searches.

The reviewers wrote to all of the authors of any RCTs and CTs that contained no data or insufficient data for the reviewers to calculate effect sizes to request such data.

On completion of the searches the citations identified were imported into the reference manager EndNote and de-duplicated (AR).

2   Criteria for including and excluding studies

Papers included in the scoping review had to be evaluations of interventions aimed at increasing the literacy or numeracy in study populations of adults. Papers were included if they were one of the following study types: randomised controlled trial (RCT), controlled trial (CT) or review of RCTs and/or CTs. Studies of this type were included on the basis that randomised and other controlled trials can be used to help to answer questions about the effectiveness of interventions designed to increase aspects of literacy and numeracy. In addition, studies were included if they were undertaken in English-speaking countries and were written in the English language. [2] Finally, papers were included if they were produced (whether formally published or not) in the years 1980-2002. Studies were excluded if they were interventions aimed at increasing the literacy or numeracy of adult (ESOL) students studying in English-speaking countries. Studies were also excluded on the basis of study type if they were interventions of a pre- and post-test design, or if they were non-interventions.

---

[2] That is countries where English is spoken as a first language by a significant segment of the population - includes UK, Ireland, USA, Canada, Australia, NZ, Jamaica and other countries in the Caribbean, Gibraltar and S. Africa, but excludes India, Pakistan, HK, Singapore, Bangladesh, Malaysia and China.

Finally, studies were excluded if all or some of the participants were aged younger than 18 years.

3   Screening

Double screening of the titles and abstracts of the studies identified by the electronic searches was undertaken by two researchers working independently (CJT and JP), using the inclusion/exclusion criteria outlined in the protocol. A measure of agreement between the reviewers was calculated using Cohen's Kappa. Any disagreements about inclusion or exclusion of papers were discussed and resolved. All papers thus identified were sent for through library inter-lending.

4   Exclusion at second stage

All received papers were double screened on the basis of the full papers. Papers were then included or excluded at second stage. Any disagreements were discussed and resolved.

5   Coding

All received included papers were then coded (by Carole Torgerson) using one of the following codes: 'RCT', 'CT', 'Review'. Reviews were included in the mapping section of the report.

6   Double data extraction and quality appraisal

For the scoping review all the RCTs and CTs were data extracted (Carole Torgerson and Jill Porthouse). Studies were coded on the basis of study type, literacy and/or numeracy focus, intervention and reviewer summary of outcome. For the two systematic reviews all RCTs, and CTs where there was sufficient data for the reviewers to calculate effect sizes, were then double data extracted and quality appraised by two reviewers (Carole Torgerson and Jill Porthouse) working independently, and using pre-established standardised data extraction tables and quality appraisal tables based on the CONSORT guidelines. The two reviewers extracted data on the development and content of the intervention evaluated, the design and results of the RCT or CT, and data on the methodological quality of the RCT. In RCTs or CTs where there was insufficient data for the effect sizes to be calculated the reviewers wrote to the authors and requested further data.

For each paper, data were extracted about: the identification and aims of the study; study design and content; use of allocation (random or otherwise) to the different groups; participants; pre- and post-intervention data; attrition rate. Full agreement was established through discussion. In addition, both reviewers independently calculated effect sizes for all studies using the computer programme Arcus Quickstat (see Appendix J). Full agreement was established. If either of the two reviewers thought that an RCT or a CT should be excluded at this third stage, agreement to exclude or retain was achieved through discussion.

7   Follow-up assessment

For the systematic review of RCTs in trials that included a follow-up assessment of learners at some point after an intervention in order to judge the extent to which gains while in provision are sustained afterwards (e.g. post-test 1, post-test 2) a comparison was made between the effect sizes at the different post-tests.

8   Wider benefits

For the systematic review of RCTs qualitative data were also extracted from any included RCT according to the following indicators: attitude to instructional method and/or testing, transfer of skills learned to other contexts, self-reported gains in literacy and numeracy, enhanced self-confidence or self-esteem, gaining employment or increase in earnings, achievement of personal goals and (for parents in family learning) greater involvement with their children's learning. These data were extracted by a researcher (Maxine Burton), who also quality appraised the design, conduct and reporting of the wider benefits of learning.

9   Syntheses

Synthesis for the scoping review took the form of a qualitative overview of all the included studies with a reviewer summary of outcome for each trial. Syntheses for the two systematic reviews took the form of a qualitative overview; a qualitative description of the main findings of the trials in terms of literacy and numeracy; and a description of the methodological strengths and weaknesses of the trials. Where it was appropriate to conduct a meta-analysis to synthesise data from two or more studies because they were homogeneous in terms of intervention and setting, this was undertaken.

10 Publication bias

It is important to assess whether the results of our review are affected by publication bias. We looked at this issue in two ways. First, we plotted the effect sizes of identified randomised controlled trials or controlled trials against their sample size (funnel plot). Second, we compared the mean effect size of published RCTs with unpublished RCTs. Undertaking a funnel plot enabled us to assess whether there was a systematic publication bias.

Comparing the means of published trials with unpublished studies enabled us to ascertain whether positive studies are more likely to be published than negative studies. If this is shown to be happening then the review will give an inflated estimate of the effectiveness of adult literacy and numeracy.

11 Comparison of effect sizes of RCTs and CTs.

Evidence from meta-analyses in educational research suggests that the effect sizes of RCTs and controlled trials are very similar, with a greater standard deviation in the controlled trials (Lipsey and Wilson, 1993). Therefore we calculated the effect sizes of the RCTs and compared them with the effect sizes of the CTs.

Appendix D

Search logs and strategies

Databases searched

ERIC via Ovid; search date 30/1/02; date coverage 1985 to November 2001.

Psycinfo via Silverplatter; search date 2/2/02; date coverage 1980 to 2002/01 week 4.

Social Scisearch via Dialog; search date 14/2/02; date coverage 1980-2002/ Feb W3

Criminal Justice Abstracts via SilverPlatter; search date 9/2/02; date coverage 1980 to 2001/12

SIGLE (System for Information on Grey Literature in Europe) via SilverPlatter; search date 9/2/02; date coverage 1980-2001/12.

SPECTR (The Campbell Collaboration's Social, Psychological, Educational, and Criminological Trials Register) via htpp://130.91.176.103/RIS/RISWEB.ISA; search date 14/2/02; date coverage not stated

Search strategies used

ERIC
1   meta analys$.mp.
2   metaanalys$.mp.
3   cochrane.mp
4   (review adj literature).ti.
5   (synthes$ adj literature).mp.
6   (synthes$ adj research).mp.
7   (synthes$ adj studies).mp.
8   (synthes$ adj data).mp.
9   (pooled adj analysis).mp
10   studies.ab.
11   (data adj pool$).ab.
12   10 and 11
13   (medline or medlars or embase or cinahl or scisearch or psychinfo or psycinfo or psychlit or psyclit).mp.
14   (hand or manual$ or database$ or computer$) adj search$).mp.
15   (review or overview).ti.
16   (systematic$ or methodologic$ or quantitative$ or research$ or literature or studies or trial$ or effective$).ti.
17   15 and 16
18   1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 12 or 13 or 14 or 17
20   (case$ adj review$).mp.
21   (record$ adj review$).mp.
22   (patient$ adj review$).mp.

23   (patient$ adj chart$).mp.
24   (peer adj review$).mp.
25   (chart$ adj review$).mp.
26   (case adj report).mp.
27   (case adj stud$).mp.
28   (prospective adj stud$).mp.
29   (book$ adj review$).mp.
30   (movie$ adj review$).mp.
31   (film$ adj review$).mp.
32   (software adj review$).mp.
33   19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32
34   18 not 33
35   evaluative.ti,ab.
36   evaluation.ti,ab,sh.
37   outcome$ stud$.ti,ab.
38   output$.ti,ab.
39   quantitative.ti,ab.
40   findings.ti.
41   allocation.ab.
42   allocated.ab.
43   random$.ti,ab.
44   intervention.ti,ab.
45   comparison.ti,ab.
46   trial$.ti,ab.
47   experiment$.ti,ab.
48   control$ trial$.mp.
49   control$ stud$.mp.
50   73 and 76
78   53 or 54 or 55 or 56 or 57 or 58 or 59 or 60 or 61 or 62 or
     63 or 64 or 65 or 66 or 67 or 68 or 69 or 75 or 77
79   adult numeracy.mp.
80   exp numeracy/
81   exp adults/
82   exp adult students/
83   adult$.mp.
84   81 or 82 or 83
85   80 and 84
86   79 or 85
87   78 or 86
88   52 and 87

PsycINFO
#1   (adult* near2 literacy) in ti,ab
#2   (adult* near2 literate)in ti,ab
#3   (adult* near2 illiteracy)in ti,ab
#4   (adult* near2 illiterate)in ti,ab
#5   (adult basic education)in ti,ab
#6   abe in ti,ab
#7   (adult* read* program*)in ti,ab
#8   (family literacy)in ti,ab
#9   (workplace education) in ti,ab

#10 (workplace literacy) in ti,ab

#11 (work near2 literacy) in ti,ab

#12 (worksite education) in ti,ab

#13 (worksite literacy) in ti,ab

#14 (community literacy) in ti,ab

#15 (adult basic skills) in ti,ab

#16 ((evenstart) in ti,ab) or ((even start) in ti,ab)

#17 explode 'Literacy-' in DE

#18 explode 'Literacy-Programs' in DE

#19 adult*

#20 #17 or #18

#21 #19 and #20

#22 literacy education and (adult*)

#23 writing skills and (adult*)

#24 reading skills and (adult*)

#25 (adult* near2 numeracy)in ti,ab

#26 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #21 or #22 or #23 or #24 or #25

Social SciSearch

1   ADULT?(2W)LITERACY

2   ADULT?(2W)LITERATE

3   ADULT?(2W)ILLITERACY

4   ADULT?(2W)ILLITERATE

5   ADULT(W)BASIC(W)EDUCATION

6   ADULT?(W)READ?(W)PROGRAM?

7   FAMILY(W)LITERACY

8   WORKPLACE(W)EDUCATION

9   WORKPLACE(W)LITERACY

10   WORK(2W)LITERACY

11   WORKSITE(W)EDUCATION

12   WORKSITE(W)LITERACY

13   COMMUNITY(W)LITERACY

14   ADULT(W)BASIC(W)SKILLS

15   (EVEN(W)START) OR EVENSTART

16   (LITERACY(W)EDUCATION) AND ADULT?

17   (LITERACY(W)PROGRAM?) AND ADULT?

18   NUMERACY(3W)ADULT?

19   (READ?(W)PROGRAM?) AND ADULT?

20   (WRITING(W)PROGRAM?) AND ADULT?

21   (READ?(W)SKILL?) AND ADULT?

22   (WRITING(W)SKILL?) AND ADULT?

23   (BASIC(W)EDUCATION) AND ADULT?

24   S1:S23

Criminal Justice Abstracts

#1   evenstart or (even start)

#2   abe

#3   literacy

#4   (correction* near2 education*) in ti,ab,de

#5   (prison* near2 education) in ti,ab,de

#6   (compensat* near2 education) in ti,ab,de
#7   writing near2 (program* or skill*)
#8   read* near2 (program* or skill*)
#9   basic education
#10 basic skills
#11 numeracy
#12 education* program*
#13 illiteracy
#14 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 or #13

SIGLE
#1   adult* near2 literacy
#2   adult* near2 literate
#3   adult* near2 illiteracy
#4   adult* near2 illiteracy
#5   adult basic education
#6   adult* read* program*
#7   family literacy
#8   workplace education
#9   workplace literacy
#10 work near2 literacy
#11 worksite education
#12 worksite literacy
#13 community literacy
#14 adult basic skills
#15 (even start) or evenstart
#16 (literacy education) and adult*
#17 (literacy program*) and adult*
#18 numeracy near3 adult*
#19 (read* program*) and adult*
#20 (writing program*) and adult*
#21 (read* skill*) and adult*
#22 (writing skill*) and adult*
#23 (basic education) and adult*
#24 abe
#25 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or
      #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24

SPECTR
1   literate
2   literacy
3   illiterate
4   illiteracy
5   adult basic education
6   (reading skills) or (reading program)
7   (writing skills) or (writing program)
8   numeracy
1 or 2 or 3 or 4 or 5 or 6 or 7 or 8

Appendix E

Characteristics of included studies-RCTs

Author, year, etc.	Study details (I = intervention group; C = control group)	Outcome
Batchelder and Rachal, 2000a USA ■ **Publication type** Journal article ■ **Setting** Maximum security prison ■ **Objective** To examine the efficacy of using computer-assisted instruction (CAI) with inmates participating in a prison education program compared with inmates participating in a traditional instruction programme using an experimental design. ■ **Study topic** Literacy and numeracy CAI Incarcerated population ■ **Outcome measures** Comprehensive Adult Student Assessment System (CASAS) maths and reading post-tests. ■ **Design** RCT (individual), digit table.	■ **Participants** n = 75 male inmates in maximum security prison. 2 ethnic groups: African-American inmates (n = 56) and Caucasians (n = 15). ■ **Intervention** I: participants received GED instructional material for 1 hour per day on computers for a total of 80 hours over a 4 week period in mathematics or language. Also traditional instruction for 3 hours per day in English, maths, history and science. C: participants received traditional instruction in English, maths, history and science for 4 hours per day for a total of 80 hours over a 4 week period.	■ **Results** Achievement scores of inmates in the intervention group were not significantly higher than those in the control group. GROUP / MEAN / SD / N CASAS math post-test Group 1: Experimental 221.9 12.3 36 Group 2: Control 217.0 17.9 35 CASAS reading post-test Group 1: Experimental 227.4 13.5 36 Group 2: Control 223.4 17.5 35 ■ **Effect size** As reported by authors: *CASAS maths* No significant difference Unadjusted effect size = 0.32 *CASAS reading* No significant difference Unadjusted effect size = 0.26 As calculated by reviewers: *CASAS maths* = 0.32 (−0.15 to 0.79) *CASAS reading* = 0.26 (−0.21 to 0.72) ■ **Reviewer summary** No difference between I and C. ■ **Comments** Study also reported as: Batchelder, 2000b attrition n = 4

Author, year, etc.	Study details	Outcome
Bean and Wilson, 1989 USA ■ **Publication type** Journal article ■ **Setting** Not stated. ■ **Objective** To examine the use of closed captioned television as a reading medium for sight vocabulary development for adults receiving literacy instruction. ■ **Study topic** Literacy (reading) ■ **Outcome measures** Sight vocabulary recognition – Word Recognition Test. ■ **Design** RCT (individual; stratified random assignment) pilot study.	■ **Participants** n = 27 (23 black, 1 white; 15 male, 8 female) students attending adult literacy program. ■ **Intervention** Ii: Closed captioned with instruction; n = 9; 5 lessons 30 minutes each over 3 week period. Iii: Closed captioned without instruction; n = 9; 5 lessons 30 minutes each over 3 week period. C: Script with instruction (traditional print instruction); n = 9; 5 lessons 30 minutes each over 3 week period.	■ **Results** There was a positive effect for the intervention (Ii: closed captioned [CC] with instruction), but this was not statistically significant. CC with instruction (n = 9) 26.4 (2.7) Script (n = 8) 24.5 (2.2) CC without instruction (n = 7) 24.9 (4.5) ■ **Effect size** As calculated by reviewers: 0.77 (−0.24 to 1.74) ■ **Reviewer summary** No difference between Ii and C. ■ **Comments** Attrition: 3 (Iii – 1; C – 2)

Author, year, etc.	Study details	Outcome
Cheek and Lindsey, 1994 USA ■ **Publication type** Journal article ■ **Setting** Urban vocational technical institute ■ **Objective** To determine the effects of two methods of reading instruction on urban adults' word identification and reading comprehension abilities (diagnostic prescriptive and traditional programmed). ■ **Study topic** Literacy ■ **Outcome measures** Stanford Diagnostic Reading Test Form B Brown Level – word identification and reading comprehension subtest scores. ■ **Design** RCT (individual)	■ **Participants** n = 76 (42 females, 29 males) economically and academically disadvantaged adults. ■ **Intervention** I: n = 38 diagnostic prescriptive approach – one hour per day for two months. Involves formal and informal diagnostic procedures to identify adults' strengths and needs. Used to develop individual educational prescriptions, ie, teaching strategies, methods and materials. C: n = 38 traditional programmed approach – one hour per day for two months. Determined present reading abilities, plus used computers and other teaching machines at their level. Focused on systematic presentation of materials and immediate feedback of responses. Each lesson was presented in step-by-step manner leading toward accomplishment of a specific goal.	■ **Results** Adults taught by the intervention method had significantly higher comprehension than adults taught by the control method but word identification was similar. MEAN SD Experimental: 50.0 0.68 Total Comprehension (n = 33) Control: 44.97 11.45 Total Comprehension (n = 38) ■ **Effect size** As calculated by reviewers (unadjusted): Total comprehension 0.46 (-0.02 to 0.92) ■ **Reviewer summary** Positive effect for I ■ **Comments** Attrition: 5 (intervention group)

Author, year, etc.	Study details	Outcome
Martinson and Friedlander, 1994 USA ■ **Publication type** Research Report ■ **Setting** California, 6 counties, adult schools and community colleges. ■ **Objective** To report findings of California's state-wide Greater Avenues for Independence (GAIN) program. To determine the effects of adult basic education classes as part of a welfare-to-work program on adult literacy skills. ■ **Study topic** Literacy ■ **Outcome measures** Educational achievement; increase in literacy as gauged by scores on a literacy test – the Test of Applied Literacy Skills (TALS). ■ **Design** RCT (individual)	■ **Participants** n = approx 1,875 (total sample size) drawn from 2,500 (from 33,000 originally randomised, 21,000 deemed to require basic education) Æ 1,719 Æ 1,119 Æ 1,115. Welfare recipients meeting GAIN's criteria for needing basic education. ■ **Intervention** I: n = 595 welfare recipients given access to GAIN's services and subject to its participation mandate, basic education services, except San Diego where efforts were made to redesign services. Mix of interventions, for example, whole class and small group teaching, CAI, individualised programmes, peer-based instruction, published workbooks, teacher-student development materials. No consistent use of employment-related examples. Average time = 10 hours per week. Average duration = one school year. C: n = 520 welfare recipients not eligible for the program and not subject to its participation mandate, but who could participate in other services in the community on their own.	■ **Results** Virtually no program impact on basic skill levels for all counties as a group. I: n = 595 mean = 475 s.d. = 94.74 C: n = 520 mean = 473 s.d. = 94.74 ■ **Effect size** As reported by authors: 0.019 As calculated by reviewers: 0.0211 -0.0965 to 0.1388 ■ **Reviewer summary** No difference between I and C. ■ **Comments** High attrition. Randomisation problems. Also published in Riccio et al. 1994, "GAIN: Benefits, costs, and Three-Year Impacts of a Welfare-to-Work Program. California's Greater Avenues for Independence Program".

Author, year, etc.	Study details	Outcome
McKane and Greene, 1996 USA ■ **Publication type** Journal article ■ **Setting** 3 minimum security (male) 1 minimum security (female) 1 minimum/maximum security (female) ■ **Objective** To examine the effectiveness of theory-based CAI compared with traditional methods of reading instruction for reading with incarcerated adults reading below the ninth grade level. ■ **Study topic** Literacy (reading) Incarcerated adults (male and female) ■ **Outcome measures** Test for Adult Basic Education post-test ■ **Design** RCT (individual)	■ **Participants** n = 150 volunteers in conventional centres who tested below the 9.1 reading level in the TABE. ■ **Intervention** I: n = 75 (?) - number unclear in original. CAI instructional group; access to computer-assisted reading instruction in addition to the traditional methods of instruction. Approximate time of reading instruction – 40 hours. C: n = 75 (?)- number unclear in original. Non-CAI instructional group; access to the traditional methods of reading instruction only; participation in CAI math instruction. Approximate time of reading instruction – 40 hours.	■ **Results** There was a small positive effect, but this was not statistically significant. CAI mean 1.59 s.d. 1.86 n = 51 Control mean 1.20 s.d. 2.02 n = 43 ■ **Effect size** As calculated by reviewers: 0.20 (-0.21 to 0.61) ■ **Reviewer summary** No difference between I and C. ■ **Comments** Attrition: huge, n = 56 (over one-third of original number randomised). Attrition due to early release. Only complete data on 94. One centre completely withdrew.

Author, year, etc.	Study details	Outcome
Nicol and Anderson, 2000 UK ■ **Publication type** Journal article ■ **Setting** Adult training centre ■ **Objective** To evaluate an experiment that compared computer-assisted and teacher-implemented instruction in numeracy (with a population of adults with 'a mild learning disability') with a third (control) group. ■ **Study topic** Numeracy Adults with 'mild learning disabilities' CAI ■ **Outcome measures** Numeracy: two psychometric tests of numeracy (basic number diagnostic test, Gillham, 1980; basic number Screening test, Gillham and Hesse, 1976). ■ **Design** RCT (individual)	■ **Participants** 'Adults with mild learning disability' n = 24 (12 males, 12 females) ■ **Intervention** Ii: CAI. One 1-hour teaching session per week for 3 months, (drill and practice and games) repeated for 3 further months. Pair work using BBC microcomputers. Iii: Conventional teaching. One 1-hour teaching session per week for 3 months, (drill and practice) repeated for 3 further months. C: Control. Structured activities that did not directly teach numeracy, eg, work or art classes.	■ **Results** Overall the 3 groups improved their numeracy scores. The teacher-led and computer groups improved more as a function of time on the intervention than did the control group. There was a positive effect for CAI compared with conventional teaching, but this was not statistically significant. |	CAI (n=8)	Teacher (n=8)	Control (n=8)
Post-test 1	111.13 (24.16)	101.88 (15.76)	98.86 (11.26)
Post-test 2	117.75 (24.83)	106.25 (17.57)	98.50 (11.03)	 ■ **Effect size** As calculated by reviewers: Post-test 1 0.46 (-0.55 to 1.44) Post-test 2 0.53 (-0.47 to 1.52) ■ **Reviewer summary** No difference between Ii and Iii ■ **Comments** Attrition: not stated.

Author, year, etc.	Study details	Outcome
Rich and Shepherd, 1993 USA ■ **Publication type** Journal article ■ **Setting** Adult Education Programs (New York) ■ **Objective** To investigate the effectiveness of a 'Reciprocal Teaching' (text comprehension) programme to improve the comprehension of poor adult readers (with sufficient cognitive ability to comprehend text). ■ **Study topic** Literacy (reading comprehension) ■ **Outcome measures** Scores on a multiple choice/short answer comprehension test and a free recall comprehension test. ■ **Design** RCT (individual) stratified random assignment (type of school attended as stratum).	■ **Participants** n = 90 students (65 female, 25 male) enrolled in adult education programs. ■ **Intervention** Ii: n = 18. Self-questioning and summarising instruction (total condition). Iii: n = 18. Self-questioning instruction. Iiii: n = 18 Summarising instruction. Small groups, 6 x 45 minutes. Lessons spaced over 18-day period. Ci: n = 18. Tests and experimental materials without instruction. Cii: n = 18. Tests only.	■ **Results** On the question task adults who self-questioned and summarised (full condition), and adults who self-questioned, significantly outperformed the two controls. *Question task* <table><tr><td>FULL</td><td>QUESTION</td><td>SUMMARY</td><td>CONTROL A</td><td>CONTROL B</td></tr><tr><td>14.0 (2.34)</td><td>13.1 (2.60)</td><td>12.4 (2.97)</td><td>10.4 (2.68)</td><td>10.6 (3.15)</td></tr></table>n = 18 for all groups *Recall task* <table><tr><td>FULL</td><td>QUESTION</td><td>SUMMARY</td><td>CONTROL A</td><td>CONTROL B</td></tr><tr><td>14.0 (6.31)</td><td>11.5 (6.83)</td><td>11.5 (7.76)</td><td>7.1 (4.59)</td><td>7.5 (3.96)</td></tr></table>n = 18 for all groups ■ **Effect size** As reported by authors: *Question task:* Full (Ii) compared to material only (CA) = 1.22 Full (Ii) compared to no intervention (CB) = 1.14 Self-questioning (Iii) compared to material only (CA) = 0.92 Self-questioning (Iii) compared to no intervention (CB) = 0.85 As calculated by reviewers: *Question task:* Ii v CA: 1.43 (0.69 2.16) Ii v CB: 1.23 (0.50 1.93) Iii v CA: 1.02 (0.32 1.71) Iii v CB: 0.87 (0.17 1.55) ■ **Reviewer summary** Positive effect for Ii and Iii

Author, year, etc.	Study details	Outcome

Schrum, 1985 USA
 	
■ **Publication type**
Research report.
■ **Setting**
Medium security prison.
■ **Objective**
To evaluate the effectiveness of a modified instructional strategy for accelerative and mastery learning of basic arithmetic by a male incarcerated population.
■ **Study topic**
Numeracy
'Incarcerated population (male)'
■ **Outcome measures**
Mastery of basic arithmetical computations as measured by Wide Range Achievement Test, Arithmetic, Level II (1978)
■ **Design**
RCT (individual)
	■ **Participants**
	n = 72 incarcerated males who had been referred for 'remediation'. All socio-economic and educational levels, all ethnic and crime groups.
	■ **Intervention**
	I: n = 12, n = 12, n = 12. Modified comprehension learning strategy system. 6 weeks preliminary, presentation and practice phases. I involved 3 phases:
	Phase 1: taught to relax, calm mind, restimulate previous pleasant learning experiences.
	Phase 2: assisted in reorganising, integrating and elaborating of previous work.
	Phase 3: practised for mastery through self-graded quizzes, spatial reorganisation of new material and review cycle for increasing recall and retention.
	C: n = 12, n = 12, n = 12. 6 weeks preliminary, presentation and practice phases.
	C conventional methods of teaching involved 3 phases:
	Phase 1: pre-test to determine current level, plus informed of topics.
	Phase 2: material presented in conventional lecture method (blackboard).
	Phase 3: completed workbook, exercises, given tests.
		■ **Results**
		The treatment resulted in higher post-test and retention test scores.
		
		Test 1: mean of treatment = 31 (7.79) n = 36
		       mean of control = 26 (7.79) n = 36
		Test 2: mean of treatment = 31 (7.79) n = 36
		       mean of control = 25 (7.79) n = 36
		■ **Effect size**
		As calculated by reviewers:
		
		Post-test 1
		0.64 (0.17 to 1.11)
		Post-test 2
		0.77 (0.29 to 1.25)
		■ **Reviewer summary**
		Positive effect for I

Author, year etc.	Study details	Outcome
St Pierre et al. 1993 **USA** ■ **Publication type** Research report ■ **Setting** Various Even Start settings, e.g., local community colleges, homes, etc. ■ **Objective** To evaluate the 'in-depth study' component of the 'Even Start' Programme; to evaluate the short- and long-term effects of 'Even Start' re. The literacy skills of adults. ■ **Study topic** Literacy and numeracy ■ **Outcome measures** Functional literacy levels on a reading test – CASAS (Comprehensive Adult Student Assessment System) ■ **Design** RCT	■ **Participants** n = 199 'Even Start' families. (I: n = 101Æ84; C: n = 98Æ75). Mixed ethnicity. ■ **Intervention** I: n = 84. Variety of activities that included adult education; services that develop the basic education and literacy skills of the adult including adult basic education (ABE), adult secondary education (ASE), English as a second language (ESL), or preparation to attain a GED certificate. C: n = 75.	■ **Results** CASAS <u>Reading Survey</u> <u>Even Start (n=84)</u> <u>Control (n=75)</u> Post-test 229.3 (13.0) 224.9 (16.7) ■ **Effect size** As reported by authors (adjusted) 0.7 (0.05 s.d.) p←0.05 As calculated by reviewers 0.29 C1 –0.02 0.61 ns ■ **Reviewer summary** Small positive effect for I, but in reviewers' judgement not statistically significant ■ **Comments** High attrition. Some controls attended adult education.

Appendix F

Characteristics of included studies-CTs

Author, year, etc.	Study details (I = intervention group; C = control group)	Outcome
Askov et al. 1986 **USA** ■ **Publication type** Research report ■ **Setting** Prison – male inmates ■ **Objective** To evaluate CAI program on literacy levels of adult male inmates compared to traditional methods. ■ **Study topic** CAI Prison inmates Literacy ■ **Outcome measures** 1) Baltimore County Design (words in isolation and words in a functional setting) 2) Bader Reading and Language Inventory 3) Slossen Oral Reading Test 4) Wide Range Achievement Test (WRAT) ■ **Design** CT	■ **Participants** n = 27 Æ n = 23 males inmates at the 0-4 reading level. n by group unknown. ■ **Intervention** I: CAI for 6 weeks, 1 hour per day. C: Small group and tutorial instruction in traditional materials for ABE. No exposure to CAI.	■ **Results** Intervention group showed statistically significant positive gains over control group on Bader Reading and Language Inventory Test. Other measures showed growth, but not statistically significant. PRE-TEST POST-TEST Intervention 0.08 0.92 Control 0.40 0.30 ■ **Effect size** Not possible for reviewers to calculate effect size. ■ **Reviewer summary** Unclear. ■ **Comments** Only reported outcome measure (mean, etc) was for test that showed statistically significant difference, i.e. Bader.
Author, year, etc.	Study details	Outcome
Broughton, 1994 **USA** ■ **Publication type** Dissertation/Thesis ■ **Setting** Job Corps Centre ■ **Objective** To assess the effectiveness of sensory-bridging computer hardware and software on academic achievements of students with sensory impairments or learning difficulties. ■ **Study topic** Sensory-bridging computer hardware and software Literacy and numeracy. ■ **Outcome measures** TABE (Test of Adult Basic Education) ■ **Design** CT	■ **Participants** n = 12 2 visually impaired, 2 deaf, 2 dyslexic in each group. ■ **Intervention** I: computer hardware/software performs functions of human surrogate readers, writers and interpreters, placed in a 'Learning Lab'. Four months duration. Example – ability to type spoken word without manipulation of a keyboard, synthetic speech capable of vocalizing text scanned from printed material. C: standard Job Corps classroom.	■ **Results** Computer group scored substantially higher than control group. No effect sizes or p scores (raw data). ■ **Effect size** Not possible for reviewers to calculate effect size. ■ **Reviewer summary** Unclear. ■ **Comments** No statistical analysis applied to results No attrition

Author, year, etc.	Study details	Outcome
Broussard, 1983 **USA** ■ **Publication type** Research report ■ **Setting** Parish school system's part-time adult education program (in elementary schools) ■ **Objective** To assess the effect of CAI on adult literacy and numeracy versus traditional classroom approach. ■ **Study topic** CAI and CMI (computer-managed instruction) Literacy and numeracy ■ **Outcome measures** Grade equivalence from California Achievement Test ■ **Design** CT	■ **Participants** No details provided. ■ **Intervention** I: n = 19 or 17 (unclear). CAI using ABE curriculum. 5 hours per week over two evenings for classes and half hour per evening of CAI for 8 months of project, plus home-based CAI (limited component at Phase 3) over 4 months, individualised. C: n = not known. Traditional ABE classes in the Parish system (and State system).	■ **Results** CAI group, both classroom and home, showed greater positive grade changes than control group (not significant). *Results (mean grade elevated score, ie, gain)* CAI (n = 17) 1.7 Control (Parish) 1.4 Control (State) 1.7 ■ **Effect size** Not possible for reviewers to calculate effect size. ■ **Reviewer summary** ■ **Comments** No detail about participants. Control group 'n' unknown.

Author, year, etc.	Study details	Outcome
Burtoff, 1985 **USA** ■ **Publication type** Research report ■ **Setting** Community Centres ■ **Objective** To investigate whether or not subjects who receive prior native language literacy instruction develop greater proficiency in the second language (English) than those who receive only second language instruction. ■ **Study topic** Literacy ESL ■ **Outcome measures** ESL proficiency: Basic English Skills Test (reading and writing) ■ **Design** CT	■ **Participants** n = 65 Ii: n = 21 Iii: n = 8 (control) ■ **Intervention** Ii: Haitian Creole literacy – 12 weeks of Haitian Creole literacy instruction followed by 12 weeks of ESL instruction. 3 times per week for 2 hours. Iii: ESL only – 24 weeks of ESL instruction. 3 times per week for 2 hours.	■ **Results** An examination of the component scores for (English) reading and writing reveals that the Creole literacy group gained more points than the ESL only group – results approach statistical significance.

Mean scores for Burtoff, 1985:

	Ii (n = 21) Pre	Ii (n = 21) Post	Iii (n = 8) Pre	Iii (n = 8) Post
Reading and writing	2.0	4.5	4.4	5.4 $(p \to 0.07)$

■ **Effect size**
Not possible for reviewers to calculate effect size (no standard deviations).

■ **Reviewer summary**
Unclear.

■ **Comments**
Attrition very high: n = 36.
No standard deviations for pre- and post-test.

Author, year, etc.	Study details	Outcome
Culclasure, 1982 **USA** ■ **Publication type** Research report ■ **Setting** State hospital State school ■ **Objective** To evaluate the use of low-cost microcomputers to improve the effectiveness of ABE services provided for 'mentally handicapped adults'. ■ **Study topic** CAI Learning disabilities Literacy (reading comprehension) ■ **Outcome measures** California Achievement Test (adapted) ■ **Design** CT	■ **Participants** n = 36 'mentally handicapped adults' ■ **Intervention** I: n = 12 received basic literacy skill (reading comprehension) instruction using low-cost microcomputer I: n = 12 received basic literacy skill (reading comprehension) instruction using programmed instruction C: n = 12 received basic literacy skill (reading comprehension) instruction using traditional classroom teaching approaches	■ **Results** CAI was superior as an instructional methodology to both programmed instruction and traditional classroom teaching. CAI mean = 75.5 Programmed mean = 70.85 Traditional mean = 54.3 p ← 0.01 ■ **Effect size** Not possible for reviewers to calculate effect size (no standard deviations, only raw data). ■ **Reviewer summary** Unclear.

Author, year, etc.	Study details	Outcome
Diem and Fairweather, 1980 **USA** ■ **Publication type** Journal article ■ **Setting** Adult Detention Center ■ **Objective** To analyse the effectiveness of a computer-assisted educational system (adult basic skills curriculum). ■ **Study topic** Literacy (reading) Numeracy CAI Prison setting ■ **Outcome measures** Adult Basic Learning Examination (ABLE) Level II ■ **Design** CT	■ **Participants** n = 38 male inmates. ■ **Intervention** I: n = 19 (n = 15). Experimental group instructed using PLATO (interactive computer graphics terminal model) in vocabulary skills, reading, spelling, arithmetic computation and arithmetic problem. C: n = 19 (n = 15). Traditional group instructed using traditional classwork instructional approach, including lecture, note recitation and team teaching in some academic areas.	■ **Results** No evidence of benefit for CAI in anything. *Table 1: Mean Achievement Gains (tenths of grade level)* | Subject Area	PLATO Group	Traditional Group
Vocabulary	1.20 (3.23)	-0.27 (1.83)
Reading	1.13 (1.99)	0.46 (5.53)
Spelling	-0.66 (1.43)	0.73 (2.70)
Arithmetic Computation	1.26 (7.40)	0.33 (3.22)
Arithmetic Problem Solving	0.84 (7.50)	1.93 (3.93)

Arithmetic Total	1.13 (6.00)	1.01 (6.22) |
		
		■ **Effect size**
		As calculated by reviewers:
		Vocabulary 0.56 (-0.17 to 1.29)
		Reading 0.16 (-0.56 to 0.88)
		Spelling -0.64 (-1.37 to 0.10)
		Arithmetic Computation
		                      0.16 (-0.56 to 0.88)
		Arithmetic Problem Solving
		                      -0.18 (-0.90 to 0.54)
		Arithmetic Total 0.02 (-0.70 to 0.73)
		■ **Reviewer summary**
		No difference.
		■ **Comments**
		Attrition: n = 8

Author, year, etc.	Study details	Outcome
Dietrich, 1994 **USA** ■ **Publication type** Research report ■ **Setting** Community college ■ **Objective** To investigate the effectiveness of auditory perception training on the reading ability of adult poor readers at a community college. ■ **Study topic** Literacy (reading) ■ **Outcome measures** Peabody picture vocabulary test-revised degrees of reading power, Form F3. Lindamood auditory conceptualisation test. Word identification subtest of the Woodcock Johnson achievement test. Word attack subtest of the Woodcock Johnson achievement test. Spelling test. ■ **Design** CT	■ **Participants** n = 30 (n = 21) adults registered for a reading and study skills course. Varied ethnicity, mixed gender. ■ **Intervention** I: Experimental group using a phonological skills approach. C: Control group using traditional metacognitive approach.	■ **Results** According to the author, the findings suggest that phonological skills do appear to be important to the reading process and that it is possible to teach phonological skills. Only variables with sufficient data were used to evaluate the effectiveness of training (LAC, word identification and word attack). The experimental group performed significantly higher than the control group on the LAC test and on Word Attack. Pre-test and post-test means (sd) of phonological measures

		Control Group			Experimental Group		
		LAC	Word Attack	Word Ident.	LAC	Word Attack	Word Ident.
Pre-test		m=71.40 sd=17.02	m=22.56 sd= 4.03	m=50.10 sd= 2.48	m=64.18 sd=21.38	m=16.45 sd= 3.60	m=46.09 sd= 3.56
Post-test		m=77.70 sd=18.30	m=23.00 sd= 2.92	m=50.20 sd= 2.20	m=89.64 sd= 8.58	m=20.18 sd= 5.46	m=47.73 sd= 2.57

■ **Effect size**
As calculated by reviewers (unadjusted):
LAC   0.85 (-0.06 to 1.74)
Word Attack   -0.63 (-1.51 to 0.25)
Word Identification   -1.03 (-1.93 to –0.10)

■ **Reviewer summary**
No difference (LAC and Word Attack)
Positive effect for control (Word Identification)

■ **Comments**
Attrition: n = 9.
Missing data.

Author, year, etc.	Study details	Outcome
Dirk and Crawford, 1993 **USA** ■ **Publication type** Journal article ■ **Setting** All-male, maximum security, state correctional facility ■ **Objective** To develop and assess an experimental module which utilised the natural world as a context for teaching reading to incarcerated adults with low reading skills. ■ **Study topic** Literacy (reading) Prison setting ■ **Outcome measures** Reading levels (T-NAT) ■ **Design** CT	■ **Participants** n = 18 male prisoners. Reading level no higher than 5th grade. ■ **Intervention** I: n = 9 'contextual learning'; a series of lessons which were designed around specific science topics of interest to the learners in the ABE program. C: n = 9 'traditional approach'; individual instruction and one-on-one teaching using traditional ABE curricular approaches and materials.	■ **Results** Both the experimental and control groups showed modest gains in reading levels. I: mean gain = 0.4 C: mean gain = 0.2 ■ **Effect size** Not possible for reviewers to calculate effect size (no standard deviations). ■ **Reviewer summary** Unclear (insufficient data). ■ **Comments** No standard deviations of pre- and post-tests given. I and C not equivalent at baseline (pre-test scores).

Author, year, etc.	Study details	Outcome
Gretes and Green, 1994 **USA** ■ **Publication type** Journal article ■ **Setting** Seven community colleges, North Carolina ■ **Objective** To determine the effect on reading skills of adults of computer-assisted instruction versus traditional reading instruction program. ■ **Study topic** CAI Literacy (reading) ■ **Outcome measures** TABE ('Total Reading' portion only) ■ **Design** CT (volunteers at three colleges randomised to intervention or control; volunteers at four other colleges assigned to control. Control participants randomly selected to match intervention group on basis of ethnicity and gender)	■ **Participants** n = 488 originally (I = 238, C = 250). Student volunteers in ABE. Ethnic mix of Caucasians, Black, Hispanic and Asians. Mixed gender. ■ **Intervention** I: READY (Reading to Educate And Develop Yourself), comprising 10 multimedia reading instruction modules covering consumer, health and citizenship issues. Delivered on CD-ROM over 11-week period. Total 44-66 hours. C: Traditional ABE reading instruction using workbooks, low-level/high-interest reading materials and conventional classroom instruction over an 11-week period.	■ **Results** CAI showed statistically significant gains over control group. Intervention group- average grade equivalent score gain of 1.8 school years. Control group- average grade equivalent score gain of 0.59 school years. p ← 0.0001. Grade equivalent gains for TABE total reading scores: MEAN SD Intervention (n = 238) 1.84 1.46 Control (n = 250) 0.59 1.10 ■ **Effect size** As calculated by reviewers: 0.9704 95% CI : 0.7821 to 1.1576 ■ **Reviewer summary** Large positive effect for I. ■ **Comments** Control numbers unclear for results. Randomisation violated – baseline equivalence at pre-test not clear.

Author, year, etc.	Study details	Outcome
Indiana OIC of America State Council, 1990 **USA** ■ **Publication type** Research report ■ **Setting** Adult basic education classes at two sites – in Indianapolis ■ **Objective** To measure the effectiveness of traditional classroom instruction versus computer-assisted instruction (CAI) in raising the competency levels of adults one grade level for each 80 hours of instruction. ■ **Study topic** CAI ■ **Outcome measures** ABLE test ■ **Design** CT	■ **Participants** n = 149 Æ n = 50 completed pre- and post-tests and attended more than 30 hours. Economically disadvantaged adults, predominantly female, aged 17 – 67 (mean age = 32.3 years) who tested below 12th grade/GED competency levels on the Adult Basic Learning Evaluation (ABLE) test. Un- or under-employed. ■ Intervention I: CAI computer-assisted teaching- no details given. n = 14 C: non-CAI traditional classroom teaching- no details given. n = 36.	■ **Results** Overall average grade change for CAI students was 2.6 grades compared with an average of 1.84 grades for non-CAI students. ■ **Effect size** Not possible for reviewers to calculate effect size (no standard deviations). ■ **Reviewer summary** Unclear. ■ **Comments** Huge attrition: 149 _ 50

Author, year, etc.	Study details	Outcome

Irby et al. 1992
USA

■ **Publication type**
Research report

■ **Setting**
Run by Centre for Adult Basic Education and Literacy Unit of Joliet Junior College. Variety of settings for ABE classes; main sites used were Fairmont Junior High School, Spanish Centre, Morris Central School.

■ **Objective**
To evaluate the effectiveness of a family literacy project on literacy and numeracy levels of adults.

■ **Study topic**
Family literacy
Literacy and numeracy

■ **Outcome measures**
TABE

■ **Design**
CT

■ **Participants**
n = unknown.
Predominantly black and Hispanic.
Results appear to relate to n = 25 (I = 15, C = 10).

■ **Intervention**
I: 'Families About Success' (FAS). Family literacy project comprising childhood component, parent child component, adult education concept component, parenting education component, career/vocational and self-esteem/self-concept components. ABE classes offered two times per week for 12 weeks. Instructors developed individualised educational plans for each student to work at his/her own pace.
C: Enrolled in GED classes only.

■ **Results**
Students in FAS showed a higher average gain in reading and math compared to controls. Data for one centre only presented (n = 25).

	Average grade level gain	
	Reading	Math
Intervention	0.91	1.21
Control	0.72	0.54

■ **Effect size**
Not possible for reviewers to calculate effect size.

■ **Reviewer summary**
Unclear.

■ **Comments**
Number of participants not clear at either recruitment or analyses.

Author, year, etc.	Study details	Outcome

Lavery et al. 1998
New Zealand

■ **Publication type**
Journal article

■ **Setting**
Government-funded employment training programme

■ **Objective**
To compare the learning outcomes associated with basic literacy education programmes conducted via traditional instruction with computer-assisted instruction (CAI).

■ **Study topic**
Literacy
Numeracy
CAI

■ **Outcome measures**
Burt Word Reading Test
Neale Analysis of Reading Ability
KeyMath Revised Test

■ **Design**
CT

■ **Participants**
n = 12, mean age 33 years, male and female, mixed ethnicity.
Adults not in paid employment.

■ **Intervention**
I: n = 6 CAI group. 18 one-hour sessions of computer-assisted instruction over period of 7 weeks. Interpretive comprehension, literal comprehension, word meaning, word analysis and reference skills; basic maths processes and their applications and mathematics problem-solving involving word problems.

C: n = 6. Traditional instruction group – pursued similar literacy and numeracy skills via a 'textbook and lecture' method over the same period.

■ **Results**
According to authors:
Significantly greater achievements were made in reading (word recognition, word accuracy and comprehension) and numeracy (mathematics concepts, operations and applications) under CAI than under traditional instruction.

Table 1: Mean scores on the Burt, Neale Accuracy and Neale Comprehension Tests at pre-test and post-test for the two teaching groups.

Tests		Traditional		Computer	
		Pre	Post	Pre	Post
Burt	M	95.33	98.17	59.00	80.17
	SD	10.63	8.75	10.79	18.61
Neale Accuracy	M	84.67	84.17	62.67	74.83
	SD	8.59	15.11	23.51	23.19
Neale C'hension	M	20.50	20.67	20.50	25.00
	SD	5.89	6.59	8.22	6.99

Table 2: Mean scores on the concepts, operations and applications subscales of the KeyMath Test at pre-test and post-test for the two teaching groups.

Subscales		Traditional		Computer	
		Pre	Post	Pre	Post
Concepts	M	42.83	43.00	40.50	48.50
	SD	7.17	10.35	9.73	7.23
Operations	M	57.33	51.83	56.50	61.33
	SD	11.52	11.27	13.55	13.03
Applications	M	64.17	64.17	59.00	73.33
	SD	14.39	15.00	13.67	11.52
Total	M	164.33	159.00	156.00	183.17
	SD	29.64	33.93	30.98	28.57

■ **Effect size**
As calculated by reviewers:
Burt	-1.24 (-2.46 to 0.04)
Neale Accuracy	-0.48 (-1.62 to 0.68)
Neale C'hension	0.64 (-0.54 to 1.79)
Math Total	0.77 (-0.43 to 1.93)

■ **Reviewer summary**
Contrary to authors' claims: no difference.

Author, year, etc.	Study details	Outcome

Lehigh County Community College, 1993USA

■ **Publication type**
Research report

■ **Setting**
Adult Basic Education Program

■ **Objective**
To compare the effectiveness of a traditional GED curriculum with a literacy curriculum based on applied literacy skills.

■ **Study topic**
Literacy

■ **Outcome measures**
Educational Testing Service's Tests of Applied Literacy Skills (TALS).
ABLE tests.

■ **Design**
CT

■ **Participants**
n = 69 (n = 28) adults participating in Adult Basic Education Program.

■ **Intervention**
I: n = 14 received GED instruction emphasising functional and workplace context and supplemental instruction. 12 hours per month for a total of 39 class sessions.
C: n = 14 received instruction based on a traditional GED curriculum. 12 hours per month for a total of 40 class sessions.

■ **Results**
No major differences between the experimental and control group members' test performance.

Pre- to post-test mean gains

	I (n = 13)	C (n = 9)
Document Literacy	12	13
Prose Literacy	40	12
Quantitative Literacy (maths) 3		20
	I (n = 10)	C (n = 9)
ABLE (reading)	1.84	- 0.1
ABLE (mathematics)	3.03	2.75

■ **Effect size**
Not possible for reviewers to calculate effect size (no standard deviations).

■ **Reviewer summary**
Unclear.

■ **Comments**
Huge attrition.
Also missing data in results.

Author, year, etc.	Study details	Outcome

Maclay and Askov, 1988 USA

■ **Publication type**
Journal article

■ **Setting**
Intermediate units

■ **Objective**
To assess effectiveness of CAI program versus no teaching

■ **Study topic**
CAI
Literacy (reading)

■ **Outcome measures**
1. Slosson Oral Reading Test (SORT)
2. Two sections of Baltimore County Design (BCD)
3. Bader Reading and Language Inventory

■ **Design**
CT

■ **Participants**
I: n = 92 Æ n = 52
C: n= 24
Parents of children receiving Chapter 1 (federally funded remedial programs in reading and math for low-income areas). Less than 4th grade reading level. Mean age: 35.6 years. Made up of Caucasian 68%, Black 21%, English 1st language 75%.

■ **Intervention**
I: CAI using a 'whole word' approach with some word building activities in teaching 1,000 high frequency and functional words. The goal is expanded word recognition. Courseware is interactive and responsive to users' answers and needs. Teachers encouraged to use courseware 80% of instructional time and to supplement with additional activities. Instructional time 20 hours in total.
C: No teaching (pre- and post-test only).

■ **Results**
Significant differences between pre-and post-test scores for intervention group and no significant change for control group.

SORT	N	Pre/M	SD	post/M	SD
Intervention	52	3.26	2.00	3.93	2.05
Control	24	2.74	1.92	2.67	1.85
BCD-E2					
Intervention	52	18.48	5.73	22.33	3.85
Control	24	17.13	5.04	16.88	4.79
BCD-E3 (words in a functional setting)					
Intervention	52	20.13	5.35	22.67	3.15
Control	24	16.00	6.22	16.13	5.95
Bader					
Intervention	52	2.33	1.80	3.63	1.76
Control	24	2.04	1.73	1.79	1.91

■ **Effect size**
As calculated by reviewers:
SORT   0.63 (0.14 to 1.13)
BCD-E2   1.31 (0.78 to 1.83)
BCD-E3   1.55 (1.0 to 2.09)
Bader   1.02 (0.50 to 1.52)

■ **Reviewer summary**
Positive effect for I.

■ **Comments**
Attrition: 40 out of 92 in intervention group did not complete.

Author, year, etc.	Study details	Outcome
Macmurdo, 1988 **USA** ■ **Publication type** Research report ■ **Setting** Various: Adult literacy centre; library; homes; church halls ■ **Objective** To determine whether the use of computers along with 'Laubach Way to Reading' materials could significantly raise the reading level of adults presently reading at 0.4 grade levels. ■ **Study topic** Literacy (reading) CAI ■ **Outcome measures** Grade levels ■ **Design** CT	■ **Participants** n = 22 ■ **Intervention** I: n = 11, working one-to-one with volunteer tutors using the 'Laubach Way to Reading' materials, twice a week for 1.5-2 hours per class, plus CAI 2-5 hours weekly for a total of 50 hours. C: n = 11, working one-to-one with volunteer tutors using the 'Laubach Way to Reading' materials, twice a week for 1.5-–2 hours per class for a total of 50 hours.	■ **Results** I: Advancement of 1.2 grades per 50 hours of work. C: Advancement of 0.7 grades per 50 hours of work. ■ **Effect size** Not possible for reviewers to calculate effect size (no standard deviations). ■ **Reviewer summary** Unclear. ■ **Comments** Original sample size n = 30. 8 'drop-outs' (4 in each group).

Author, year, etc.	Study details	Outcome
Meyer et al. 1983 **USA** ■ **Publication type** Research report ■ **Setting** Prison; six institutions in Illinois Department of Corrections school district. One maximum and two medium security in each group. ■ **Objective** To evaluate the effectiveness of a CMI program on literacy and numeracy of adult male inmates, compared to traditional program. ■ **Study topic** CMI Prison ■ **Outcome measures** TABE ■ **Design** CT	■ **Participants** n = 359 adult male inmates. 70% minority groups. Average age 25 years. ■ **Intervention** I: n = 207. PLATO computer-managed program, 2.5 hours per day over 3 months. n reduced to 139 at post-test for reading/math and n = 83 for language. C: 153. Traditional, self-paced program, 2.5 hours per day over 3 months. n reduced to 53 at post-test for reading/math and n = 11 for language.	■ **Results** Results showed significant gains for ALL students in language, math and reading. See table below. ■ **Effect size** As calculated by reviewers: Reading 0.2235 (-0.0939 to 0.5403) Math 0.2466 (-0.071 to 0.5636) Language 0.2111 (-0.4187 to 0.8407) ■ **Reviewer summary** No difference. ■ **Comments** Attrition 53% due to turnover in medium security prisons. (66% from control group, 33% intervention).

	Pre-test	s.d.	Post-test	s.d.
Reading				
Intervention (n = 139)	7.6	1.46	8.5	1.73
Control (n = 53)	7.2	1.55	8.1	1.94
Math				
Intervention (n = 139)	7.0	1.03	8.0	1.63
Control (n = 53)	7.1	1.15	7.6	1.60
Language				
Intervention (n = 83)	6.3	1.45	8.3	1.72
Control (n = 11)	5.5	1.69	7.9	2.96

Author, year, etc.	Study details	Outcome
Morrow et al. 1993 **USA** ■ **Publication type** Research report ■ **Setting** ABE classes run by Greater Pittsburgh Literacy Council in Alleghney County. ■ **Objective** To evaluate small-group instruction program on reading levels of adults compared to traditional one-to-one instruction. ■ **Study topic** Reading Teaching strategy – small group instruction ■ **Outcome measures** ABLE (Adult Basic Learning Examination) – reading comprehension. SORT (Slosson Oral Reading Test) – word recognition. ■ **Design** CT	■ **Participants** I: n = 20 Æ n = 9, mostly men. C: n = 11, approx equal men and women. White and black equal in both groups. ■ **Intervention** I: Small-group instruction, 1.5 hours twice-weekly over 7-month period. Average hours = 46. Staffed by a teacher and an aide. Learner-generated issues incorporated into curriculum discussion. Students encouraged to read and write as often as possible. Teachers became facilitators who shared authority and planned and provided instruction to meet individual student needs. Learners were consulted about their preferences with regard to materials, methods, etc. C: Traditional one-to-one instruction. Students 'matched' to tutor by area co-ordinator who also provided prescriptions for appropriate learning procedures and materials. Progress evaluation after each 50 hours of instruction. Average hours = 63.	■ **Results** Both groups made positive gains in reading comprehension but did not differ significantly across the two groups. No gains in word recognition. *Reading comprehension (ABLE test) mean scores:* |	Pre-test	Post-test	
---	---	---	---
Intervention	604.33	621.33	(p = 0.125)
Control	638.20	656.30	(p = 0.052) | *Word Recognition (SORT test) mean scores:* |	Pre-test	Post-test
Intervention	4.6	5.4
Control	3.8	4.45 | ■ **Effect size** Not possible for reviewers to calculate effect size. ■ **Reviewer summary** Unclear. ■ **Comments** Intervention group attrition rate: 47.06% (n = 20 Æ n = 9)

Author, year, etc.	Study details	Outcome
Nurss, 1989 **USA** ■ **Publication type** Research report ■ **Setting** Adult Education Centre ■ **Objective** To assess the effectiveness of PALS program on literacy of adult non-readers compared to traditional non-computerised program. ■ **Study topic** CAI Literacy ■ **Outcome measures** TABE ■ **Design** CT	**Participants** n = 209 (99 female, 110 male). ■ **Intervention** I: n = 135 participants received PALS (IBM Principles of Adult Literacy Program). Classes held daily; students attended 1-4 times per week over 8 months. C: n = 74 participants received traditional non-computerised program.	■ **Results** According to the author, there were no significant differences in post-test scores between groups (p ← 0.16). |	Pre-test Mean	SD	Post-test Mean	SD
---	---	---	---	---
Intervention (n = 42)	2.52	1.15	2.86	1.31
Control (n = 11)	4.84	1.78	5.55	2.59	 ■ **Effect size** As calculated by reviewers: –1.6386 (–2.3678 to –0.8961) ■ **Reviewer summary** Contrary to the author's statement, significant positive effect for control. ■ **Comments** Huge attrition.

Author, year, etc.	Study details	Outcome
Rio Salado Community College, 1991 **USA** ■ **Publication type** Report ■ **Setting** Factory/work environment ■ **Objective** To determine if home-based CAI, when used as a supplement to traditional ABE classes, affects reading levels. Called SALSA (Southwest Advanced Learning System for Adults). ■ **Study topic** CAI – home-based Literacy ■ **Outcome measures** TABE, Form 6 ■ **Design** CT	■ **Participants** n = 60 (30 per group) production line factory workers, between 4th and 12th reading grades. ■ **Intervention** I: Home-based CAI in addition to traditional ABE classes in work-based environment. Given training how to use, and telephone assistance available. No further details. C: Traditional ABE classes.	■ **Results** CAI group showed larger positive gains than control group, but difference was not statistically significant. ■ **Effect size** Not possible for reviewers to calculate effect size. ■ **Reviewer summary** Unclear. ■ **Comments** Originally had 53 matched pairs, which were abandoned. Reason stated was due to attrition in control group, but later stated as due to high pre-test scores.

Author, year, etc.	Study details	Outcome
Roberts et al. 1994 **USA** ■ **Publication type** Journal article ■ **Setting** Medium security prison ■ **Objective** To test the effects of an innovative group-process intervention technique on reading performance among a population of incarcerated adult males in a medium security prison. ■ **Study topic** Literacy Prison setting ■ **Outcome measures** Gates-MacGinitie Reading Test ■ **Design** CT	■ **Participants** n = 91 male prisoners, mean age 29.9 83% black, 17% white ■ **Intervention** I: Experimental Group (Community Building Group Process and SRA reading program). Two and a half day intensive community building workshop, followed by weekly ongoing sessions. Attendance was required for first two weeks, after which attendance at the programme became voluntary. SRA reading program: based on a model of co-operative learning involving teacher instruction, team practice, individual assessments and group recognition. n = 31 Ci: Control Group 1 (SRA reading program only). n = 34 Cii: Control Group 2 (untreated). n = 26	■ **Results** Mean G-M-gain scores were significantly greater for the intervention than the two control groups. GE Gain Mean ESS Gain Mean I 1.04 (1.15) p = 0.0001 23.93 (24.82) p = 0.0001 Ci -0.14 (1.00) -0.62 (24.41) Cii -0.01 (1.02) -5.31 (25.66) ■ **Effect size** I compared with Ci as calculated by reviewers: GE: 1.1 (0.57 to 1.62) ESS: 1.0 (0.48 to 1.51) I compared with Cii as calculated by reviewers: GE: 0.96 (0.41 to 1.51) ESS: 1.16 (0.59 to 1.72) ■ **Reviewer summary** Positive effect for I.

Author, year, etc.	Study details	Outcome
Schrader, 1984 **USA** ■ **Publication type** Research report ■ **Setting** Technical Institute ■ **Objective** To determine the value and effectiveness of a developmental remedial reading program. ■ **Study topic** Literacy (reading) ■ **Outcome measures** Grade point average ■ **Design** CT	■ **Participants** n = 44 full-time technical institute students with reading skill deficiencies. gender = not stated. ■ **Intervention** I: n = 20 participated (voluntarily) in the developmental remedial reading program (instruction and activities in comprehension, vocabulary and decoding); small group and individual instruction with reading specialist, a variety of techniques, equipment and materials. C: n = 24 did not participate in program.	**Results** No significant difference between the means of grade point averages in I and C. I: 2.18 (1.58) n = 20 C: 2.28 (1.58) n = 24 ■ **Effect size** As calculated by reviewers: -0.063 (-0.66 to 0.53) ■ **Reviewer summary** No difference. ■ **Comments** 'Assumed' equivalence at baseline.

Author, year, etc.	Study details	Outcome
Smith and Dalheim, 1990 **USA** ■ **Publication type** Research report ■ **Setting** ABE Center, two local churches and Pittsfield Salvation Army. ■ **Objective** To improve the reading levels of diagnosed 'learning disabled adults' ('Project READ'). ■ **Study topic** Literacy (reading) ■ **Outcome measures** Oral reading (accuracy and comprehension) Word attack Oral reading comprehension Silent reading comprehension ■ **Design** CT	■ **Participants** n = 20 'learning disabled' adults, ranging in age from young to retired. ■ **Intervention** Ii: 'reading from scratch' – a structured, sequential, synthetic phonics curriculum. Iii: 'reading from scratch with equipment' (eg, tape recorder). C: traditional adult reading approach, focusing on language experience. All groups participated for entire programme 1 hour per day, 4 days per week for 2 years.	■ **Results** The direct teaching of phonics and language structure in a controlled, sequential format is more effective for learning disabled readers than is teaching in which the inclusion of phonics instruction is at the discretion of the teacher. Oral Reading and Word Attack mean gains _Ii (n = 8) C (n = 4)_ Oral Reading Accuracy gains 2.6 1.9 Comprehension gains 4.2 1.3 Word Attack 3.0 1.8 Oral and Silent Reading Comprehension Gains _I (n = 8) C (n = 4)_ Oral reading comprehension 4.2 1.3 Silent reading comprehension 3.6 0.2 ■ **Effect size** Not possible for reviewers to calculate effect size (no standard deviations). ■ **Reviewer summary** Unclear. ■ **Comments** Group composition changed after 6 months. High attrition.

Author, year, etc.	Study details	Outcome
Thuy, 1992 **USA** ■ **Publication type** Research report ■ **Setting** Indochinese American Council (IAC) ESL Programs. ■ **Objective** To study the educational effectiveness of computer-assisted instruction (CAI) in a combined English as a Second Language (ESL) and Family Literacy curriculum. ■ **Study topic** CAI ESL Literacy ■ **Outcome measures** Vocabulary and comprehension questions taken from the readings. ■ **Design** CT	■ **Participants** n = 30; 15 in intervention and control, but two experimental groups (numbers not stated). Intervention group – mixture of Vietnamese and Chinese. Control group– mainly Cambodians. Ii: Beginning/higher beginning level group. Iii: Beginning level group C: Beginning level group. ■ **Intervention** I: CAI (four software packages). C: Printed material only. Each group received 6 hours instruction per week. For the intervention group, 2 hours were on materials in software.	**Results** Both groups improved on post-test. Intervention group showed bigger gains than control group.

Results (Thuy, 1992)

	Pre A*	Post A	Pre B**	Post B
Paper Bag Princess				
Intervention (i)	56.67	86.25	78.00	84.50
Intervention (ii)	3.19	67.00	46.00	84.25
Control	3.13	17.33	9.13	27.80
Thomas' Snowsuit				
Intervention (i)	67.72	86.44	76.00	86.78
Intervention (ii)	16.23	49.47	20.33	51.47
Control	3.00	13.63	12.17	33.33
Classified Ads				
Intervention (i)	29.33	45.33	8.50	9.00
Intervention (ii)	3.40	40.40	0.00	9.50
Control	5.20	12.73	2.87	5.00
Christmas Season				
Intervention (i)	25.50	35.00	9.00	9.50
Intervention (ii)	11.71	36.36	1.86	8.86
Control	8.42	15.29	3.83	6.25

*Test A = Vocabulary; **Test B = Comprehension*

■ **Effect size**
Not possible for reviewers to calculate effect size.

■ **Reviewer summary**
Unclear.

Author, year, etc.	Study details	Outcome
Wadsworth and Frazier, 1982 **USA** ■ **Publication type** Research report ■ **Setting** City Skill Centre and Opportunities Industrialisation Centre. ■ **Objective** To evaluate the effect of the PLATO instructional system on adult literacy and numeracy. ■ **Study topic** Literacy and numeracy CAI ■ **Outcome measures** Grade level achievement ■ **Design** CT	■ **Participants** n = 244 at Skill Centre n = 71 at OIC ■ **Intervention** Group 1 – had less than 20 hours of PLATO in a given subject. Group 2 – had more than 20 hours of PLATO in a given subject.	■ **Results** No significant differences between groups on any of the outcomes of language, reading and maths.

Results (Wadsworth and Frazier, 1982)

	Pre-test	Post-test
Language		
Controls (n = 41)	6.2	9.2
Less than 20 hrs (n = 10)	5.9	8.6
More than 20 hrs (n = 8)	5.4	7.8
Reading		
Controls (n = 21)	6.5	7.9
Less than 20 hrs (n = 18)	6.3	8.1
More than 20 hrs (n = 18)	6.3	7.7
Mathematics		
Controls (n = 65)	6.0	8.8
Less than 20 hrs (n = 66)	6.1	8.4
More than 20 hrs (n = 29)	5.6	8.6

■ **Effect size**
Not possible for reviewers to calculate effect size.

■ **Reviewer summary**
Unclear.

■ **Comments**
Attrition: Skill Centre n = 73 (30%)
                OIC n = 15 (21%)

Author, year, etc.	Study details	Outcome
Washington (US Dept of Education), 1991 **USA** ■ **Publication type** Research report ■ **Setting** Community college ■ **Objective** To evaluate the effectiveness of a CAI program on basic adult literacy, compared to traditional classes. ■ **Study topic** CAI Literacy ■ **Outcome measures** Assessment test developed by the college. ■ **Design** CT	■ **Participants** n = 42 (20 intervention, 22 control), predominantly women, aged 25-40 years. n reduced to 23 completers. ■ **Intervention** I: n = 9 received CAI with audio support plus traditional literacy classes. Students had usual classes and CAI was available to them in their own time. Work was individualised. 18 – 20 tutorial lessons and associated audio tapes, topic-based. C: n = 14 received traditional literacy classes, individualised instruction. 'Developmental English Classes'.	■ **Results** Average gain presented only. Average change score Intervention (n = 9) 2.85 Control (n = 14) 1.60 ■ **Effect size** Not possible for reviewers to calculate effect size. ■ **Reviewer summary** Unclear. ■ **Comments** Attrition: n = 42 Æ 23.

Author, year, etc.	Study details	Outcome
Wilson, 1992 **USA** ■ **Publication type** Report ■ **Setting** Community College ■ **Objective** To determine the value of a computer-assisted approach on adult literacy and numeracy compared to more traditional teaching approaches. ■ **Study topic** CAI Literacy and numeracy ■ **Outcome measures** TABE literacy and numeracy (Control 2) Canadian Adult Achievement Test (Control 1) ■ **Design** CT (historical controls)	■ **Participants** I: n = 15 (9 female, 6 male) mean age: 32.27 years mean educational level: 9.33 years Ci: n = 10 (7 male, 3 female) mean age 20.3 years mean educational level: 10.11 years Cii: n = 15 (8 male, 7 female) Mean age: 38.4 years Mean educational level: 8.71 years ■ **Intervention** I: n = 15 received CAI for 11 weeks. Prior to formal beginning students had two full days of orientation to the computer and software. Work focused on reading v mathematics. Weekly schedule determined in conjunction with instructor in week 1 followed by 11 weeks training.	■ **Results** I: significant positive change in problem solving (p ← 0.005). Ci: significant positive change in reading comprehension and vocabulary (p ← 0.05). Ci and Cii: significant positive change in number operations and maths (p ← 0.001). ■ **Effect size** Not possible for reviewers to calculate effect size. ■ **Reviewer summary** Unclear. ■ **Comments** In comparison with Cii, some subjects excluded from analysis as no data available on the measure used (TABE), therefore results may not be reliable as n = 5.

Author, year, etc.	Study details	Outcome
Wisher and O'Hara, 1981 **USA** ■ **Publication type** Research report ■ **Setting** Navy ■ **Objective** To compare the standard classroom approach to 'academic remedial training' with the use of computer-assisted instruction (CAI). ■ **Study topic** CAI Literacy (reading) ■ **Outcome measures** Gates–MacGinitie reading test (comprehension section) Navy recruit reading test (NRRT) ■ **Design** CT	■ **Participants** n = 152 ■ **Intervention** I: n = 77 (n = 47 restricted sample) received academic remedial training through CAI C: n = 75 (n = 47 restricted sample) received academic remedial training through classroom instruction	■ **Results** No overall significant difference in general reading comprehension skills between the groups. I C NRRT Pre-test 17.6 17.0 Score Post-test 21.1 20.1 Gates- Pre-test 40.1 39.7 MacGinitie Post-test 43.3 42.2 ■ **Effect size** Not possible for reviewers to calculate effect size (no standard deviations). ■ **Reviewer summary** Unclear.

Appendix G

Quality of included trials (RCTs and CTs)

In this Appendix we consider the quality of each of the 18 trials for which an effect size could be calculated.

Randomised controlled trials

Batchelder and Rachal (2000)
The aim of this study was to examine the effectiveness of using Computer-Assisted Instruction (CAI) for literacy and numeracy compared with using traditional instruction. Seventy-five male inmates in a maximum security prison (mean age 30.5 years) were randomly assigned to receive either GED instructional material in maths and language on computers and traditional instruction in English, maths, history and science (intervention group) or to receive traditional instruction in English, maths, history and science (control group).

The authors used a random number table to assign individuals to their groups. The total number that appeared to be randomised at the outset was 75, although only 71 were included in the study. The authors did not state what happened to the missing 4 participants, which suggests that intention to teach analysis was not undertaken (Torgerson and Torgerson, 2003). The authors appeared to have used simple randomisation. For a group size of 71 this would probably have resulted in a numerical imbalance between the groups, which did not happen (i.e. there were 36 in the experimental group and 35 in the control group). The randomisation process did not appear to have been concealed from the investigators. The authors did not report whether or not the administration and marking of follow-up tests was undertaken 'blindly'. There were no statistically significant differences at post-test. In maths an effect size of 0.32 was calculated (CI –0.15 to 0.79); in reading an effect size of 0.26 was calculated (CI –0.21 to 0.72). However, the upper 95 per cent confidence intervals for both maths and reading did not exclude 0.7 of a standardised effect, which is a large difference. The study could therefore have experienced a Type II error i.e. finding no 'statistically significant' benefit, when there could actually have been an educationally significant benefit of the intervention.

Bean and Wilson (1989)
The aim of this study was to examine the effectiveness of close captioned television as a reading medium. Tweny-seven students attending an adult literacy program were randomly assigned to the close captioned mode with instruction, to the close captioned mode without instruction or to a script mode with traditional print instruction.

In this trial stratified random allocation was used, resulting in three equally sized groups of 9 students each. Three students were lost to follow-up leaving three groups of: 9, 8 and 7 respectively. Clearly this sample size is low and would not have had very much statistical power to detect even one standard deviation difference in improvement (which requires at least 32 participants in a two-group comparison). The method of allocation was not described, although the students were stratified on baseline reading scores, which is probably the strongest and therefore most important co-variate. The authors did not state whether or not

the post-tests were administered 'blindly'. The effect size of the intervention was considerable (point estimate 0.77, CI −0.24 to 1.74) but not statistically significant. Because of the tiny sample size, there was a high possibility of a Type II error.

Cheek and Lindsey (1994)

The aim of this study was to determine the effects of two methods of reading instruction on adults' word identification and reading comprehension abilities. Seventy-six adults were randomly assigned to a 'diagnostic prescriptive' approach or to a traditional approach to reading instruction.

In this trial 76 adults were randomly allocated to two classes. The authors stated that random assignment had occurred but that this had been controlled to ensure equal numbers of adults in each class (matched in initial reading level). The method by which this was achieved was not stated, nor was whether or not the allocation procedure was undertaken independently from the researchers. The two classes were taught by two instructors. It was not made clear if the same two instructors taught both classes, which would have controlled for inter-teacher variability, or whether 4 instructors were used: two different ones for each class. If it were the latter approach the effects of the intervention would have been perfectly confounded with the abilities of the instructors, making it impossible to ascribe changes to a treatment or instructor effect. Outcome assessment does not appear to have been conducted blindly. There were 5 drop-outs, all in the experimental group. Any form of attrition can lead to bias; however, differential attrition is of greater concern than equal attrition as this increases the possibility of attrition bias (Torgerson and Torgerson, 2003). The authors undertook six statistical comparisons in reading ability with only two being of borderline statistical significance. For total comprehension, the effect size was positive for the intervention (0.46) and of borderline statistical significance (-0.02 to 0.92).

Martinson and Friedlander (1994)

The objective of this study was to determine the effects of adult basic education classes as part of a welfare-to-work program of adult literacy classes. More than 20,000 participants took part in this study although only 1,115 were randomised and followed up with appropriate data collection. This study, although the largest that was identified in the review, had a number of methodological problems. Firstly, there was a high attrition rate of 35 per cent, which can lead to bias. Secondly, there were differences in numbers between the control and experimental groups. Six counties in California were involved in the study and centres within each county randomised participants to either the intervention group or to act as controls. In five of the six counties the percentages randomised to the experimental group ranged from 68 per cent to 86 per cent, whilst in the remaining sixth county there were approximately 50 per cent in each group (which one would expect with simple randomisation in such large groups). Five counties must have used unequal allocation; however, this was not reported in the paper. Possibly the five study centres may have formed control groups through other means, such as allowing adults who did not want to participate in the interventions to act as controls. Clearly if this happened then the results from those 5 counties are unreliable. Unfortunately, the process of random allocation was not clearly described in this paper. In terms of the intervention it is difficult to ascertain which component of the intervention actually worked, as a 'black box' approach was adopted with a package of different interventions offered differently by different sites. A better approach to try and disentangle what may have worked and what did not work would have been to use a randomised factorial design.

McKane and Greene (1996)

The objective of this study was to examine the effectiveness of theory-based computer-assisted-instruction (CAI) for reading instruction compared with traditional methods of reading instruction. In this study 150 male and female adults in maximum and minimum security prisons were randomised to computer-assisted-instruction (CAI) or to no CAI. However, there were only 94 participants available for final analysis: an attrition rate of 37 per cent. Because the study did not state the initial group allocation size it is not possible to ascertain whether differential attrition took place. It is also clear that intention to teach had not taken place as those participants with 'abnormal' gains or losses in their change scores were excluded from the analysis. The control treatment included a mixture of teaching methods which, whilst pragmatic, makes it difficult to ascertain exactly what was occurring in the control group. The analysis was problematic in that the authors broke the sample up into three 'sub-groups' and undertook a separate analysis of each. This approach increases the chances of a spurious chance finding (i.e. the more statistical tests that are done the greater the chance of finding a 'statistically' significant difference, when in truth there is no difference). The authors could have also undertaken a multivariate analysis that included baseline scores and each individual prison as co-variates in order to control for chance differences between groups at randomisation and centre effects. Adjusting for co-variates is particularly important when there are relatively small numbers in the sample. The outcome measures used by these authors may have been inappropriate as the mean score was less than two, as was the standard deviation. This suggests the post-test used may have been too difficult for the participants and would therefore have been relatively insensitive to changes. There was a small positive effect for CAI at post-test, but this was not statistically significant.

Nicol and Anderson (2000)

The objective of this study was to evaluate an experiment that compared computer-assisted and teacher-implemented instruction in numeracy. The authors undertook a very small trial of only 24 adults who had been randomised into 3 groups of 8. The method of random allocation was not described but it is implied that some form of stratified allocation on the basis of gender was undertaken to ensure balance in this co-variate, and numerical equivalence. There was no justification of the sample size. There was no difference in improvement between the teacher-led intervention and the CAI, but given the very small numbers in each group there is high possibility of a Type II error in this study. It was not clear whether or not the same teachers taught the two intervention groups. If they were different teachers, then there is likely to have been a problem of confounding between the teacher and the intervention. The groups were not well balanced at baseline, with the CAI group having much higher pre-test scores than the no CAI group (over half a standard deviation). The inclusion of a 'no teaching' control group was unnecessary as the important policy question is between conventional teaching and CAI. Therefore, it would have been a more efficient use of the available sample size to allocate the control students into the two treatment arms. The graphical display of the results in this study was misleading with the plots not reflecting the true differences by giving an inflated appearance of the difference between the groups. There was a positive effect for the CAI, but this was not statistically significant.

Rich and Shepherd (1993)

The objective of this study was to investigate the effectiveness of a 'reciprocal teaching' programme to improve text comprehension. The authors randomised 90 participants to one of 5 groups. Randomisation was stratified by centre and all group sizes were identical (n = 18). The stratification process was probably constrained to produce numerically equivalent group sizes although this is not formally stated. The authors undertook post-testing 'blind' to the

group allocation. There did not seem to be any attrition and therefore all randomised participants were included in the final analysis. The authors undertook baseline testing of co-variates, which is unnecessary and can be misleading. Statistical tests make the assumption that any difference is not due to chance whereas any differences between randomised groups at baseline must be due to chance (if the randomisation process is fair). Chance imbalance is a real threat to the analysis, especially with small sample sizes. The method of dealing with chance bias is to pre-specify co-variates in advance of the analysis and fit them in a multivariate model irrespective of the size of that imbalance (Senn, 1989). Testing for imbalance is misleading because although gender imbalance is not statistically significant in this study there is still quite a difference (approximately 20 per cent between certain groups) and if gender is an important co-variate this will still affect the results. Nevertheless, the authors did adjust for pre-test score, which is likely to be the most important co-variate. The results indicated a large and strong positive effect (statistically significant) for all of the training strategies (i.e. about 1 standard deviation) relative to the two control strategies. Interestingly, the outcomes for the control group that received the educational materials and essentially had to self-teach were similar to the outcomes for the control group that received no materials.

Schrum (1985)

The objective of this study was to evaluate the effectiveness of a 'modified' instructional strategy for accelerative and mastery learning of basic numeracy. 72 males in a medium security prison were randomly assigned to a modified comprehension learning strategy system or to conventional methods of teaching. In this study the participants were randomly allocated using random number tables; the authors did not state whether or not this was concealed. Because equal numbers were always allocated to the same groups the authors were probably not using simple random number tables as this would have been unlikely to produce numerically equal groups. The intervention group received a wide range of interventions, ranging from relaxation techniques, to classical music as well as a novel teaching style, whereas the control group received instruction using traditional didactic methods. Because the intervention group received a package of treatments it is impossible to disentangle the effects of any one component from the teaching method. The authors did not state whether or not post-tests were given 'blindly'. Baseline covariate testing was not necessary. There is a discrepancy in the analysis of post-test scores. In the analysis of variance there appeared to be 72 participants included in the analysis (i.e. all the participants) but in the univariate analysis of mean scores there were only 57 participants (i.e. 31 in the intervention and 26 in the control). There appears to have been a 20 per cent attrition rate in some of the post-tests. The treatment resulted in higher post-test and retention scores.

St. Pierre et al. (1993)

The objective of this study was to evaluate the 'in-depth study' component of the 'Even Start' Programme. In this study 199 families were randomised to an Even Start programme or to act as controls. To date the relevant papers describing the exact methodology have not yet been received. In the report containing the results section a number of methodological problems can be observed. There appeared to be some uneven attrition with the Even Start families having 84 present at post-test compared with 75 of the controls present at post-test. Whilst the groups appeared to be balanced at baseline they were unbalanced in terms of pre-test scores. The authors compared the means of the change scores, which does not fully deal with the imbalance. The authors found a small effect which was not, in the reviewers' judgement, statistically significant. This effect could be either an under- or overestimate because of the inappropriate statistical analysis that was undertaken. The authors should have undertaken a

multivariate analysis of the change scores with the pre-test scores as co-variates, which would have dealt with the imbalance.

Controlled trials

Diem and Fairweather (1980)

The objective of this study was to investigate the effectiveness of a computer-assisted educational system on literacy and numeracy. The experimental group received instruction in vocabulary skills, reading, spelling, arithmetic computation and arithmetic problem using computer-assisted-instruction. The control group received instruction in the same academic areas using a traditional class work instructional approach.

The participants were male inmates of an Adult Detention Center.

Equivalence at baseline was not established because the pre-test scores were not reported, but the mean gain scores were used in the analysis. The authors explained that the 'negative gains' and tiny gain scores in some outcomes could 'safely be treated as zero – merely artefacts of testing error.

The results indicate no evidence of benefit for CAI in anything, i.e. no statistically significant difference for each of the six outcome measures.

Dietrich (1994)

The objective of this study was to investigate the effectiveness of a phonological skills approach on the reading of 'adult poor readers' at a community college. The experimental group used a phonological skills approach to ready; the control group used a traditional metacognitive approach.

The participants were 30 students of mixed gender and varied ethnic backgrounds registered for a Reading and Study Skills course at a Community College.

Although 'two equal size groups' were randomly assigned to experimental or control condition, this was not a randomised controlled trial as there were only two groups. Attrition was not stated but 9 out of 30 students were eliminated due to incomplete data. The two groups were not equivalent at baseline. Indeed, the author states that she deliberately placed those students 'most in need of phonological help' in the experimental group, thus introducing the probability of selection bias. However, this imbalance at baseline was taken into consideration in some of the analyses.

Unadjusted effect sizes (as calculated by reviewers) showed no difference between the two groups for two outcome measures and positive effect for control for one outcome measure.

Gretes and Green (1994)

The objective of this trial was to establish the effect of CAI on the reading skills of adults that were attending seven community colleges for basic education classes. The CAI intervention (READY: Reading to Educate And Develop Yourself) comprised ten multimedia reading instruction modules covering consumer, health and citizenship issues. Instruction time totalled between 44 and 66 hours for each participant over an 11-week period. The control group received traditional adult basic education reading instruction using conventional classroom instruction over the same 11-week period.

The participants in this study were predominantly Black (53 per cent in both groups) but also included a substantial proportion of Caucasians (37 per cent in the CAI group and 38 per cent in the control group). There were also a number of Hispanics and Asians, 5 and 3 per cent respectively, in each group. All of the participants were student volunteers in adult basic education classes with slightly more women in each group than men (women 56 per cent, men 44 per cent).

In order to form the two groups the authors used a total of seven community colleges. Three of the colleges that were sufficiently equipped served as the intervention group. The control group was formed by collecting data on adults attending traditional ABE classes at all seven colleges and then matching the percentage composition of the CAI group on the basis of ethnicity and sex. Whilst these may be important covariates, the authors did not match on pre-test score, which is arguably one of the most important covariates. However, they did go on to test for equivalence between the groups and found them to have equivalent reading abilities. Despite this measure, the inherent problem with a case-control study such as this is that the groups may still differ on other unknown covariates that may influence outcome. Therefore, we cannot be completely sure that the apparent gain in reading skills was due to the intervention.

This trial demonstrated a large, positive, statistically significant effect for the intervention.

Lavery et al. (1998)

The objective of this study was to compare the effectiveness of computer-assisted-instruction and traditional instruction in literacy and numeracy. The experimental group received CAI in interpretive comprehension, literal comprehension, word meaning, word analysis and reference skills, basic maths processes and their applications and mathematics problem solving involving word problems. The control group received similar literacy and numeracy instruction via a 'textbook and lecture method'.

The participants were male and female adults of mixed ethnicity and not in paid employment.

This study used a small sample size (12). The author states that 'an intact-groups, pre-test/post-test quasi-experimental design was employed in order to assess the progress of the two groups. Random assignment of adults to groups was not possible in this study, thus the design is subject to the same threats to internal validity characteristic of much educational research.' However, two separate researchers administered pre- and post-tests, and they did not appear to be involved in the intervention. There was a lack of baseline equivalence (noted by the authors) because of lack of random assignment. However, pre-test scores were included in the analysis to control for pre-existing differences.

Interpretation of gain scores is problematic because the authors used mean gain scores, but the reviewers used unadjusted means of post-tests. According to the authors, significantly greater achievements were made in reading and numeracy under CAI than under traditional instruction. The reviewers disagree with these findings: according to our calculations, none of the effect sizes were statistically significant, and we conclude that this experiment showed no effect.

Maclay and Askov (1988)

The objective of this study was to assess the effectiveness of a CAI (CAI) program on the literacy skills of adults who had less than the 4th grade reading level. The participants were

all parents whose children were recipients of a federally funded remedial programme in reading and maths for low-income areas in Pennsylvania. The mean age was 35.6 years and the participants were predominantly Caucasian, with English the first language for 75 per cent of the participants.

Adults in the intervention group received a CAI program for an instructional time totalling 20 hours. The program was interactive, responding to users' answers and needs, and aimed at expanding recognition of 1,000 high frequency and functional words. The control group did not receive any teaching and were simply pre- and post-tested.

In this trial 92 adults were allocated to the intervention (CAI) group and 24 acted as controls. Participants were pre-tested at baseline for their reading level and then a post-test was conducted at the end of the study. The post-test (outcome assessment) was undertaken by the teacher of each class. This suggests that they must have been aware of the group allocation. 'Blinding' or 'masking' of the outcome assessor is important in order to minimise bias in the assessment. Since the teachers were not blind to the group allocation they may have been influenced by that knowledge, resulting in a biased assessment of outcome. This can lead to an over- or under estimation of the effectiveness of the intervention.

The study showed a large, positive, statistically significant effect for the CAI program compared to no teaching.

Meyer et al. (1983)

The objective of this trial was to evaluate the effects of a CAI program on adult prisoners' basic literacy and numeracy skills. Participants in the intervention group received the CAI program (PLATO – Program Logic for Automated Training Operations) for 2.5 hours per day over a 3-month period. Those in the control group had traditional adult basic education for the same amount of hours.

All the participants were adult, male inmates in either maximum or medium security prisons. Approximately 70 per cent were from ethnic minorities with a mean age of 25 years.

At the beginning of this study there were 207 participants in the CAI group and 153 in the control group. Due to turnover in the prisons many participants were lost to follow-up and not included in the analysis, resulting in 66 per cent attrition in the control group and 33 per cent in the intervention group. Any attrition can lead to bias but this possibility is increased where the attrition is differential between the two groups, as was the case in this trial, leading to the possibility of a biased outcome and an over or under estimation of the effectiveness of the intervention. Attrition bias can be minimised by continuing to gather outcome data, even though participants were no longer receiving the intervention, and including all participants in the analysis in their original group allocation (intention-to-teach analysis).

The findings from this study were that all participants showed significant gains in language, math and reading; however, there was no difference between the two groups.

Nurss (1989)

The authors of this study wished to assess the effectiveness of the PALS CAI program on the literacy skills of adult non-readers, compared to traditional adult basic education. Participants in the intervention group attended CAI classes between 1 and 4 times per week over a total period of 8 months. The control group received traditional adult basic education.

All the participants were students at an adult education centre, comprising 98 females and 109 males. 135 students were allocated to the intervention and 74 acted as controls. However, this trial experienced huge attrition of the participants; 68 per cent of the intervention and 85 per cent of the control group failed to complete the study and were excluded from the analysis. As previously noted, differential attrition between the groups means that there is the possibility of attrition bias, which may affect the outcome.

This trial showed a significant, positive effect for the traditional adult basic education classes, i.e. the control group.

Roberts et al. (1994)
The objective of this study was to investigate the effects of a 'community building group process' intervention on the reading performances of adult males.

The participants were 91 male prisoners in a medium security prison. The inmates were assigned to 'dormitories' by the correctional authorities.

Three dormitories were chosen randomly to participate, and then randomly assigned to intervention or control i or ii (untreated control). However, this was not a randomised controlled trial because there were only three groups. The authors demonstrated that the sample was representative of the overall population at the correctional facility (baseline test results revealed no significant difference among the three groups).

The results indicated a positive effect for the community building intervention at this correctional facility.

Schrader (1984)
The objective of this study was to investigate the effectiveness of a developmental remedial program. The experimental group participated in the developmental program that included instruction and activities in comprehension, vocabulary and decoding. The program used small group and individual instruction with a reading specialist using a variety of techniques, equipment and materials. The control group did not participate in the program.

The participants were full-time technical institute students with 'reading skill deficiencies'.

The author 'assumed' equivalence at baseline, but this was not demonstrated.

The results revealed no significant differences between the means of grade point averages in the intervention and control groups.

Appendix H

Tables of studies included: wider benefits

Batchelder and Rachal, 2000 USA

Source of indicators
Interviews with participants and staff.

Attitude to instructional method/testing
Dislike of multiple testing: 'negative attitude' of many even from those initially excited.
Motivation for participation not primarily educational for some (i.e. get away from cell-block).
Many intimidated by computer instruction without teacher support and felt like guinea pigs in a lab experiment.

Comments
First two stated as affecting both experimental and control groups.

Bean and Wilson, 1989 USA

Source of indicators
Attitude survey
Teacher observation

Attitude to instructional method/testing
100 per cent enjoyment of working with closed captioned TV.
Students with higher reading levels reported (by teachers) as being less positive and would prefer to read about topics from books.
25 per cent did not enjoy script lessons.

Self-reported gains in learning
For majority closed captions helped them:
1. learn new information
2. pronounce new words
3. learn meaning of new words

Comments
Not clear what form 'attitude survey' took.
Inconsistency – 100 per cent enjoyment, but higher level students (proportion not stated) reported as less positive.

Cheek and Lindsey, 1994 USA

No data

Martinson and Friedlander, 1994 USA

Source of Indicators
Survey of staff attitudes.
Survey of participants – length of time/no of hours a week of attendance.
Anecdotal (teachers/administrators/researchers).

Attitude to instructional method/testing
Staff views mainly positive on basic education as 'worthwhile activity' for participants, but with proviso on possible stigma, (see 'Confidence' below).
Suggestion that less motivation for participants to take TALS test compared with GED, which was viewed as a step towards higher earnings.
Attendance problems may be due to 'unresolved problems or negative experiences in school'.

Transfer of skills learned to other contexts
TALS test purported to measure skills relevant to employment and 'life skills' i.e. literacy outside purely academic setting. Except in one county virtually no impact and only weak correlation between number of hours in basic education and scores. TALS registered least impact at lowest literacy levels.

Confidence and/or self-esteem
Suggestion that participants may 'lack motivation, confidence and school-relevant skills' (see * in Comments below).
View stated that separate GAIN classes might isolate or stigmatise students.

Interaction with children's learning
See Comments.

Change in job, status and/or earnings
No impact on employment within 2-yr follow-up period (too short a time).
No impact on receipt of vocational training or post-secondary education but impact on use of job search services and small increase in work experience.

Achievement of personal goals
No goals stated from participants' point of view. 16 per cent completed program within 1 yr but nearly 50 per cent left without completing. Increased receipt of GED but no impact on receipt of trade certificates.

Other
Increased participation in education programs. However primary effect stated as increasing the number of adults who returned to education rather than the overall amount of time they stayed in it.

Comments
Educational attainment /achievement was stated as expected to lead to increased employment and earnings, improved well-being of participants' children etc. However also made point that 'partly owing to the relatively short follow-up period this report does not fully address the relationship between educational attainment and achievement and economic outcomes.' Questions of children's well-being also 'beyond the scope' of the GAIN evaluation.
* Emphasis placed on fact that the participants were 'under mandate' to attend so may lack qualities that are possessed by voluntary students in adult education.

McKane and Greene, 1996 USA

No data

Nicol and Anderson, 2000 UK (Scotland)

Source of indicators
Teacher observation, anecdotal.

Attitude to instructional method/testing
Enjoyment of CAI, of being 'in control' (observation).
Transfer of skills learned to other contexts
1. tendered more appropriate amounts of cash while shopping.
2. improvement of skill in placing horse-racing bets (anecdotal).

Comments
Transfer of skills not stated as being benefit just of CAI.
Proportion of students affected never stated.

Rich and Shepherd, 1993 USA

Source of Indicators
Participant comment and teacher observation?

Attitude to instructional method/testing
Instruction 'favourably regarded' by the subjects – deduced from 'positive comments' and 'active involvement in the sessions'.

Comments
Proportion of subjects not stated.
Source/prompt for indicators never stated.

Schrum, 1985 USA

Source of indicators
'Noted' by 'some of the treatment subjects'.

Transfer of skills learned to other contexts
'Felt better equipped to learn other subjects'.

Confidence/self-esteem
'Felt better about themselves'.
'Felt more in control of their lives' in sense that 'could survive on streets without resorting to crime, drugs or alcohol'.
(change of lifestyle)

Comments
Prompt for indicators unclear.
Stated as applied only to treatment groups.

St. Pierre et al. 1993 USA

Source of indicators
Interview with parents, which included:
The Inventory of Socially Supportive Behaviours (ISSB); questions on family resources – sources of income, level of income and adequacy of family resources; question on employment status (F/T, P/T, not working).
Focus groups (for views of program).

Attitude to instructional method/testing
'Several' found availability of early childhood services important – convenience and/or safety factors in being able to bring children to classes.
Staff described as 'respectful and willing to do anything for the families', treated participants 'as human beings without putting you down or judging you' while 'helping us in a new way of life'; 'friendly and supportive environment'.

Confidence and/or self-esteem
Reduced sense of isolation and of being 'trapped at home with their children'.
Felt better able to cope with life situations.
Felt valued and concerns taken seriously.
Higher personal goals and more self-esteem reported by 'a number of' participants.

Interaction with children's learning
Could communicate more effectively with their children.
Learned how to play with and plan activities for their children.
Could interact more positively, deal with stresses, 'several' had more patience, more realistic expectations.

Change in job, status and/or earnings
Slight, but not significant, increase in income.
No short-term effect on employment.

Achievement of personal goals
Most parents indicated that parenting classes were not their initial motivation for joining Even Start, but rather the adult education – preparing for GED, improving English, improving their lives were motivational factors.
Describe themselves as 'moving towards their goals of an educational certificate, getting a job and being a better parent.'

Self-reported gains in learning
'learning gains' described by participants.

Other
No effect on social support (i.e. help and support from relatives, friends and neighbours).

Comments
Hypothesis stated that Even Start would help families develop a wider social network, and greater access to social services; would enable parents to find or improve employment, with long-term effect of increasing income and reducing reliance on public assistance.
Emphasises distinction between 'qualitative changes' in participants' lives as positive short-term impacts and the 'quantifiable variables' of income and employment which require longer interventions.

Appendix I

Proposal for a randomised controlled trial of adult literacy and numeracy interventions.

1. Background

Poor adult literacy and numeracy is an important problem in English-speaking nations. A systematic review and meta-analysis has found only 9 randomised controlled trials, the majority of which were undertaken in the USA. Eight of these studies were small, and all had methodological problems. Nevertheless, they do indicate that adult literacy and numeracy classes may be beneficial. Before such classes are encouraged within the workplace they need to be evaluated within the confines of a rigorous randomised trial. This proposal describes such a study.

2. Aims and purpose of Investigation

- To assess whether work-based adult literacy and numeracy classes are effective
- To measure the costs and benefits of such classes.

3. Plan of Investigation

3.1 What is the proposed trial design?
The proposed trial is a three-armed pragmatic randomised controlled trial. Participants will be randomised, in equal numbers, to one of three groups: no intervention; literacy classes; or numeracy classes (Figure 1).

3.2 What are the planned trial interventions?
For the control group there will be no specific intervention. Adult literacy and numeracy experts will advise on the length and content of the interventions

Adults with Low Literacy and Numeracy Skills Pre-tests		
Randomised to Numeracy Classes	Randomised to Literacy Classes	Randomised to Control Group
1 month Post test	1 month Post test	1 month Post test
3 months Post test	3 months Post test	3 months Post test
6 months Post test	6 months Post test	6 months Post test
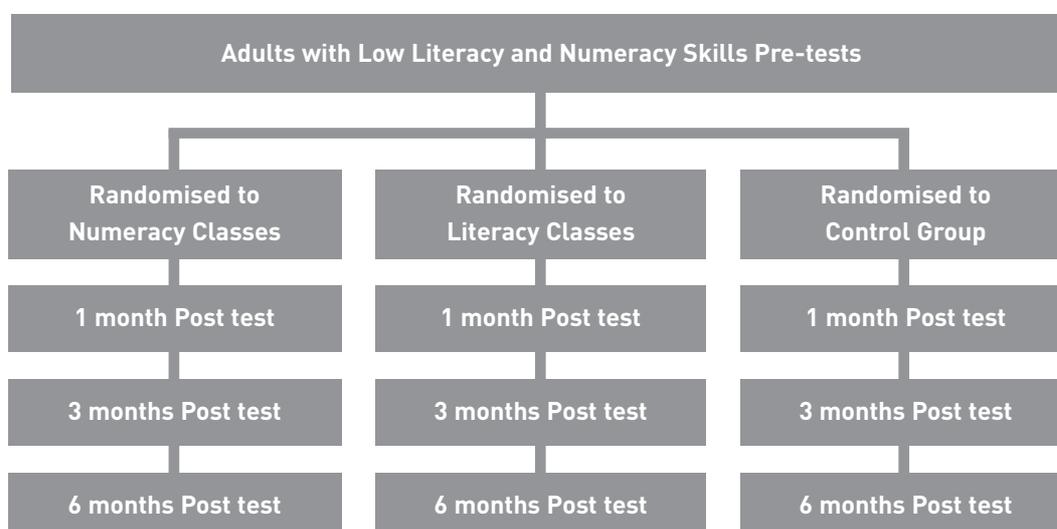
**Figure 1: Outline of trial design**

3.3 What are the planned inclusion/exclusion criteria?

Inclusion criteria

To maximise the generalisability of our trial results we will include as wide a range of adults with low levels of literacy and numeracy as possible. There will be no age limits to be included in the study. Only adults (aged 18 years and over) who perform lower than a pre-set criterion (to be determined) on the pre-tests in literacy or numeracy will be eligible. We will also include adults that are receiving educational classes outside the workplace as we wish to measure the additive effect of workplace classes.

Exclusion criteria

To minimise loss to follow-up we will exclude adults who do not anticipate being with the same employer at the end of the study (i.e. in 6 months' time). For example, adults working out their notice will be excluded.

3.4 What is the proposed frequency and duration of follow-up?

The study will last for 6 months. Participants will be given follow-up measures at 0 (i.e. pre-test), 1, 3 and 6 months after randomisation. Participants leaving the course will be asked to complete the follow-up measures.

3.5 What outcome measures will be used and what will be measured?

Primary measures

Participants will be given the same outcome measures at pre-test, before randomisation. At follow-up the participants will be invited to attend a classroom where they will be given post-tests under the supervision of a researcher or teacher who is blind to group membership. Adult literacy and numeracy experts will advise on the appropriate literacy and numeracy tests.

Secondary measures

In addition to measures of educational outcome we will include questionnaires asking participants about their satisfaction with both the educational classes and their jobs. In addition, there should be put in place a method of measuring self-esteem and confidence. Questionnaires or interviews that seek evidence on self-reported gains, whether in confidence, learning, application of skills to other contexts, or life-style changes, should not be administered to participants by the teachers directly responsible for their learning.

We will also ask whether promotion prospects have been enhanced since the classes. Finally, we will put into place a method of following up the participants long-term (e.g. 18–24 months after the last post-test) by asking them if we can re-contact them in the future. This will allow us to assess any longer term benefits of the classes.

3.6 What are the arrangements for allocating participants to the trial groups?

Random allocation will be concealed from the investigators and tutors. Participants will be individually allocated by computer, using distance randomisation. Researchers randomising the participant will ring the remote randomisation service, during working hours, and give the participant's name, age, gender and pre-test score to the randomisation service. This will be entered into the computer, which will produce a unique identifying number and the allocated intervention. The researcher will be told these over the telephone. This method will prevent any compromising of the random allocation as the computer database will 'lock in' the participants' details, preventing any group change after randomisation. The randomisation will be stratified on age (below or above 25 years), gender and pre-test score (below or above

the median). This will ensure that the three groups are well balanced on the major co-variates.

3.7 What are the proposed methods for protecting against other sources of bias?
Stratified randomisation protects against chance and selection bias. There are other forms of bias that can occur.

Hawthorne effect and 'resentful' demoralisation
Because participants cannot be blinded to their educational arm there is a threat of bias from the Hawthorne effect. This is where participants in the educational arms appear to be doing better merely because they are getting more attention from the researchers and tutors (i.e. Hawthorne effect) and any observed benefit is due to this rather than to the intervention itself. We could have addressed this by including a non-numeracy or -literacy 'placebo' group (e.g. art lessons). We feel this is unnecessary in this instance as the literacy and numeracy classes can act as controls for each other. Indeed, a placebo group could introduce an effect of its own which might promote literacy and numeracy. Because of this we feel a 'no-treatment' arm is justified because of the possibility of 'spill-over' effects of literacy on numeracy and vice-versa. A no-intervention control group would enable us to observe any spill-over. This also addresses the other threat of resentful demoralisation, which should not occur among the treatment groups.

Attrition bias
One method we propose to use to minimise attrition bias is by offering £10 to each participant when they have completed each post-test (i.e. £30 in total). This should help reduce attrition in follow-up. Participants who leave their post during the trial and move out of the area will be paid travel expenses to return for their post-test.

Observer bias
Observer or researcher bias will be minimised by having the post-tests marked blindly by two independent tutors. Further, the post-tests will be given to all groups at the same time with the test invigilator blind to participants' group status.

Analytical bias
All analysis will be undertaken by a statistician blind to the group allocation (i.e. the groups will be identified only as A, B, C).

Teacher or instructor bias
Adult classes may appear to be effective or ineffective if a significant proportion of the students were taught by a very good or a very poor teacher. To protect against this form of bias we will ensure we have at least ten teachers teaching the literacy groups and ten teaching the numeracy groups. This will reduce the possibility of one 'outlier' affecting the main results.

4. Analysis

The data analysis will be conducted using the principle of 'intention to teach'. All participants will be analysed at post-test according to their initial random allocation. The analysis will take the following form. First, we will undertake an analysis of the means between the groups adjusting for pre-test scores. Thus, the difference between pre- and post-test scores will be

the dependent variable, with the pre-test and group allocation acting as co-variates. The second analysis will fit the additional co-variates of employer, gender and age to the regression model. A statistical significance level of $p = 0.05$ will indicate statistical significance and 95 per cent confidence intervals will be calculated.

5. Secondary analyses

We will undertake further analyses to examine if there is a 'dose-response' relationship between attendance at adult literacy classes and post-test scores. This will be done by using linear regression of regressing the number of classes attended against the post-test scores. We will also explore whether there is an interaction between pre-test scores and subsequent post-test scores by group allocation. These latter analyses will be hypothesis-forming.

6. What will the sample size be and how was it derived?

Lipsey and Wilson (1993) in their meta-analyses state that most educational interventions generate an effect size of about 0.30, whilst our meta-analysis has indicated an effect size of 0.88 with the bottom 95 per cent confidence interval being 0.50. Because we are comparing our intervention against effectively no treatment we would expect a somewhat greater effect size than 0.30. We have therefore decided to power our study to observe a treatment effect of 0.5 of a standard deviation. To do this with 80 per cent power and a 5 per cent significance level would require 63 participants in each group (i.e. total number of 196). If we inflate the sample size to allow for an attrition rate of 15 per cent we will require 222 participants (i.e. 74 in each group).

7. Economic analysis

Cost data on setting up the classes will be collected as part of the study. We will then be able to describe the costs of setting up the classes and set these alongside the benefits to the students of undertaking the classes. We can estimate the benefit of the classes to the wider economy by looking at published data on the differential in wage rates between adults with low literacy and numeracy levels with those of levels that were attained by the classes (assuming a positive effect).

Appendix J

Copy of help from Arcus QuickStat for meta-analysis of continuous outcomes

Case-control studies of continuous outcomes (e.g. serum creatinine) may be investigated with respect to overall size of effect of an intervention. Meta-analysis may be used to investigate the combination or interaction of a group of independent studies, for example a series of effect sizes from similar studies conducted at different centres. This Arcus function examines the effect size within each stratum and across all of the studies/strata.

There are a number of statistical methods for estimating effect size, Arcus uses g (modified Glass with pooled sample standard deviation) and the unbiased estimator d (ref 71):

Experimental: Ne (number), Mue (mean), Se (standard deviation)
Control: Nc (number), Muc (mean), Sc (standard deviation)
J(m): bias correction factor
N=Ne+Nc
Pooled sd = sqr(((Ne-1)*Se^2+(Nc-1)*Sc^2)/(N-2))

g = (Mue-Muc)/pooled sd
d = g * J(N-2)

For each study Arcus gives g with an exact confidence interval and d with an approximate confidence interval. An iterative method based on the non-central t distribution is used to construct the confidence interval for g (ref 71).

The pooled mean effect size estimate (d+) is calculated using direct weights defined as the inverse of the variance of d for each study/stratum. An approximate confidence interval for d+ is given with a chi-square statistic and probability of this pooled effect size being equal to zero (ref 71).

Arcus also gives the option to base effect size calculations on weighted mean difference (a non-standardised estimate unlike g and d) as described in the Cochrane Collaboration Handbook.

The Q ("combinability") statistic is given with its associated probability on k (number of strata) minus one degrees of freedom. This has low power as a strict test of homogeneity, it is included here as a part of the DerSimonian-Laird random effects analysis. There are no comprehensive rules on when to use random effects and when to use fixed effects models; debate continues in the statistical community.

Please note that the results from Arcus may be slightly different from the results you obtain using other packages or from those quoted in papers; this is due to the use of exact bias correction calculated from the gamma distribution in Arcus.

Data input
You may enter number, mean and standard deviation for control and experimental groups of

each study. Alternatively you may just enter numbers in experimental groups, numbers in control groups and effect size g (nb. please make sure you use g only as defined above!).

EXAMPLE (personal communication from Dr N. Freemantle):
The following data represent test outcomes for six studies in which an educational intervention was investigated:

Trial	Experimental			Control		
	N	Mu	SD	N	Mu	SD
Kotte	27	18.5	14.9	17	5.4	17.3
Levinson	16	60	29.2	15	55	29.2
Oliver (intensive)	25	10.72	6.46	66	6.92	6.83
Oliver (standard)	62	9.2	6.16	66	6.92	6.83
Sulmasy	9	3.75	2.55	22	1.05	2.12
White	63	60	13.3	40	46.3	18.6
Wilson	23	9.24	5.35	23	5.33	4.48

To analyse these data in Arcus first prepare them in four worksheet columns and label these columns appropriately. Alternatively, open the test worksheet using the file open function of the file menu. Then select risk difference from the meta-analysis section of the analysis menu. Select the columns marked "Exposed total", "Exposed cases", "Non-exposed total" and "Non-exposed cases" when prompted for data. Note that "exposed" and "experimental" groups are the same.

For this example:

Stratum	J(N-2)	g	Exact 95% CI	Study
1	0.982	0.8261	0.190 to 1.453	Kotte
2	0.9739	0.1712	-0.536 to 0.8759	Levinson
3	0.9915	0.5644	0.0954 to 1.030	Oliver (intensive)
4	0.994	0.3500	-2.108 to 0.6986	Oliver (standard)
5	0.9739	1.2017	0.3581 to 2.0267	Sulmasy
6	0.9925	0.8804	0.4639 to 1.2928	White
7	0.9828	0.7924	0.187 to 1.3893	Wilson

Stratum	N (experimental)	N (control)	D	Approximate 95% CI	Study
1	27	17	0.8113	0.1812 to 1.4413	Kotte
2	16	15	0.1668	-0.5389 to 0.8724	Levinson
3	25	66	0.5597	0.0923 to 1.0271	Oliver (intensive)
4	62	66	0.3479	0.0013 to 0.6972	Oliver (standard)
5	9	22	1.1703	0.3419 to 1.9987	Sulmasy
6	63	40	0.8738	0.4600 to 1.2876	White
7	23	23	0.7788	0.1794 to 1.3783	Wilson

Pooled estimate of effect size d+ = 0.6124
Approximate 95% CI = 0.4212 to 0.8035

Chi-square (for d+) = 39.4424 (df = 1) P ← 0.0001

Q ("combinability" for d+) = 7.7377 (df = 6) P = .258

DerSimonian-Laird pooled d+ = 0.6278
Approximate 95% CI = 0.403 to 0.8525

DerSimonian-Laird chi-square = 29.9728 (df = 1) P ← 0.0001

Here we can say with 95 per cent confidence, assuming a random effects model, that the true size of the effect was at least 0.4 greater for the group who received the educational intervention compared with those who did not. Assuming a fixed effects model a slightly stronger inference could be made about an effect size of 0.42 (the lower confidence limit) but the high inter-study variation makes the fixed effects model less appropriate.

71. Hedges and Olkin, Statistical methods for meta-analysis. Academic Press, London, 1985.

www.nrdc.org.uk