Educational Studies,
Vol. 32, No. 2, June 2006, pp. 133–143

Is an intervention using computer software effective in literacy learning? A randomised controlled trial

G Brooks[a]*, JNV Miles[b], CJ Torgerson[b]* and DJ Torgerson[b]
[a]*University of Sheffield, UK;* [b]*University of York, UK*

Background: computer software is widely used to support literacy learning. There are few randomised trials to support its effectiveness. Therefore, there is an urgent need to rigorously evaluate computer software that supports literacy learning.

Methods: we undertook a pragmatic randomised controlled trial among pupils aged 11–12 within a single state comprehensive school in the North of England. The pupils were randomised to receive 10 hours of literacy learning delivered via laptop computers or to act as controls. Both groups received normal literacy learning. A pre-test and two post-tests were given in spelling and literacy. The main pre-defined outcome was improvements in spelling scores.

Results: 155 pupils were randomly allocated, 77 to the ICT group and 78 to control. Four pupils left the school before post-testing and 25 pupils did not have both pre- and post-test data. Therefore, 63 and 67 pupils were included in the main analysis for the ICT and control groups respectively. After adjusting for pre-test scores there was a slight increase in spelling scores, associated with the ICT intervention, but this was not statistically significant (0.954, 95% confidence interval (CI) – 1.83 to 3.74, p = 0.50). For reading scores there was a statistically significant decrease associated with the ICT intervention (–2.33, 95% CI –0.96 to –3.71, p = 0.001).

Conclusions: we found no evidence of a statistically significant benefit on spelling outcomes using a computer program for literacy learning. For reading there seemed to be a reduction in reading scores associated with the use of the program. All new literacy software needs to be tested in a rigorous trial before it is used routinely in schools.

Keywords: *Randomised controlled trial; RCT; ICT; Spelling; CONSORT*

Background

There have been too few rigorous evaluations of initiatives designed to boost schoolchildren's literacy skills (Brooks, 2002), and even fewer at secondary level.

* Corresponding authors: Department of Educational Studies, University of Sheffield, Sheffield S10 2TN, UK. Email: G.Brooks@sheffield.ac.uk; Department of Educational Studies, University of York, York Y010 5DD, UK. Email: cjt3@york.ac.uk

ISSN 0305-5698 (print)/ISSN 1465-4300 (online)/06/020133–11

DOI: 10.1080/03055690500416116

One initiative widely used, but under-evaluated, is the use of information and communication technology (ICT). Although successive governments have invested heavily in ICT in the belief that this will improve educational outcomes, there is little rigorous evidence to sustain this view.

An area of literacy where ICT might have a role is in the teaching of **spelling**. A recent systematic review and meta-analysis looking at the role of ICT on spelling acquisition found only seven small randomised controlled trials (RCTs) that evaluated the effectiveness of interventions using ICT in spelling instruction (Torgerson & Elbourne, 2002). Only one study, published in 1989, actually showed a statistically significant benefit of the use of computers to teach spelling (McClurg & Kasakow, 1989). In that study of 35 US pupils a very large effect size was found (Cohen's d =1.15, 95% confidence interval (CI) 0.44 to 1.87). However, when this study was included in a meta-analysis with five other studies, the benefit of computer-aided learning of spelling was no longer statistically significant (pooled effect size = 0.37, 95% CI –0.02 to 0.77). Nevertheless there was a trend towards a modest benefit of ICT. However, the existing studies were small, with the largest only including 79 students. None of the studies had been undertaken within the United Kingdom. Furthermore, the most recent study was undertaken more than six years ago.

A more recent systematic review and meta-analysis of the effectiveness of literacy learning (5–16 years) (Torgerson & Zhu, 2003) found only two RCTs evaluating the use of computer-aided instruction in **reading** (Lin *et al.*, 1991; Heise *et al.*, 1991). One of the trials (Heise *et al.*, 1991) found a non-statistically significant moderately positive effect size for reading, and the other (Lin *et al.*, 1991) found non-statistically significant negative effect sizes for two populations of learners (children with difficulties and normally achieving children).

Given that there is little evidence from randomised controlled trials for the effectiveness of ICT on either spelling or reading outcomes, there is an urgent need, therefore, to undertake a rigorous evaluation of the use of ICT in the teaching of **spelling and reading**.

The aim of the study reported in this paper was to assess whether an intervention using software (a computer-based spelling and reading program) would boost the literacy, particularly reading and spelling, attainment of Year 7 (age 11–12) pupils.

Method

Study design

We undertook a pragmatic randomised controlled trial. The RCT is the only research method that can adequately control for all the unmeasured variables that may affect student outcomes. Randomisation ensures that all potential confounding factors are distributed without bias across the randomised groups, and controls for temporal and regression to the mean effects (Torgerson & Torgerson, 2001). Some RCTs are designed to be explanatory studies; that is, they wish to answer the question whether an intervention works in an ideal setting. For example, in a trial included in the

Torgerson and Elbourne (2002) systematic review (see above), Berninger and colleagues (1998) undertook an RCT of computer-supported spelling learning within a university environment. In this case the results were difficult to extrapolate to a 'usual' school setting.

We wanted to establish whether or not an ICT intervention is effective within a 'normal' school environment. Therefore, we adopted the pragmatic approach. All students within our study were exposed to normal classroom teaching, which included the usual ICT and English lessons. The only difference was that half were randomised to be withdrawn from usual lessons to receive an intensive course of the ICT intervention evaluated in this study. Those pupils who acted as controls for the RCT were given the intervention immediately afterwards (see below).

Setting

The study was undertaken among Year 7 pupils (aged between 11 and 12 years) at a comprehensive (secondary) school in the North of England.

Intervention

Children who were randomised to the intervention group received the program, which is designed to boost pupils' reading and spelling through phonological awareness/word attack skills. The pupils undertake dictation processes using their own recorded voice linked to text. The arrangement for the use of the software was mainly that each pupil received the program for an hour a day for 10 consecutive school days, 10 hours in all. However, because of timetabling constraints in the school one group received the program for two hours a day for five days, and another for one hour a day for eight days plus one day at two hours.

The rationale for the intervention is that learners are thought to benefit significantly from hearing themselves vocalise pieces of text with immediate feedback and the constant opportunity for self-correction. The program has various levels suited to pupils at different levels of attainment. Pupils are allocated to the levels by brief reading and spelling tests separate from those used in this evaluation. Pupils work independently and at their own pace. In keeping with the pragmatic nature of the trial the intervention was delivered by the pupils' usual teaching assistants, who had received training for the purpose from the designer of the program.

The standard teaching of reading and spelling to Year 7 pupils in this school was for those who had difficulties in literacy to be given ICT support for literacy learning. Because of the perceived 'success' of this approach in the same school using a pre- and post-test design in the school year 2003/04, it was decided to give the intervention to *all* children entering the school in Year 7 in the Autumn term of 2004. Nevertheless, the gains observed in 2003/04 may have been simply due to temporal effects or to the regression to the mean phenomenon.

The decision to give the intervention to all Year 7 entrants, rather than just to those already known to have or thought to have literacy difficulties, was based on two factors: the school has one of the two lowest-attaining intakes of all publicly funded secondary schools in its Local Education Authority area, so that an across-the-board improvement was desirable (though notice that the mean pre-test scores for the pupils in this study were close to national norms); and the number of entrants (155) gave sufficient statistical power to observe an educationally important difference (see below).

The intervention was delivered via laptop computers, and the school had a room with several study carrels where pupils could sit and use the program with minimum disruption to others. However, this did mean that only six pupils could receive the program at any one time. Because of the extra demand for the intervention there were hardware constraints and therefore pupils had to be 'staggered' in their use of the program. Some had access to the program early in the term, whilst others accessed the software later in the term. In the absence of the trial this allocation would have been arbitrary or through administrative convenience. Because it was not possible for all pupils to have access to the software at the same time this gave an excellent opportunity in which to conduct a rigorous trial as to its effectiveness.

Control

Children randomised to the waiting-list control group received the usual literacy teaching that is standard practice for the school.

Participants

Agreement was secured from the school, and in particular from the head teacher, to host this project in the Autumn term 2004. All pupils entering the school in September 2004 (n=155) took part in the evaluation; all were randomly assigned to one cohort or the other.

Random allocation

It was important that the randomisation process was carried out independently of the assessment of the pupils, to avoid the possibility of selection bias due to subversion of the randomisation process. To achieve this each child was given a unique code and the codes were sent to an independent researcher to randomly allocate them into the two groups. Because the independent researcher did not have access to information about the characteristics of the children, selection bias could not occur by, for example, the researcher putting students with high test scores selectively into one of the groups. However, this also prevented us from using stratified randomisation, which reduces the possibility of chance imbalance. However, to reduce the problem of chance imbalance we used analysis of covariance (see below; this also has the additional advantage of increasing the power of the statistical test).

Once randomisation had been carried out, all the pupils were given the pre-test, and the group allocation was only released to the school after completion of the pre-test. (Because of resource limitations, each cohort received the intervention in batches – see below.) The randomisation was undertaken by a computer program. This was done as follows. Three blocks of random numbers were produced: 60, 60 and 40 (five allocations in the last block were unused). Within each block the computer produced equivalent numbers of control and intervention allocations. It was necessary to produce three blocks to account for the staggering of the intervention.

The pupils were randomly assigned on an individual (not school-class) basis to receive the software program in two cohorts. Both cohorts received the program within the Autumn term but one soon after the other.

Sample size calculation

Our sample size was dictated by the number of children that were enrolled into the new school year. In the 2004 intake to this school there were 155 new pupils, which is our sample size. Nevertheless, it is important to consider the likely effect size our sample would be able to detect and whether this is of 'educational significance'.

A review of the impact of educational and psychological interventions (Lipsey & Wilson, 1993) has shown that, on average, effective educational interventions rarely produce effect sizes greater than half a standard deviation ($d = 0.5$). Therefore, sample sizes for educational trials should have the power to detect, as a minimum, such a difference. The sample size in this study is large enough to observe this difference. The number of pupils included in the study would give in excess of 80% power to observe as significant (2-tailed $p = 0.05$) a moderate effect size of 0.5 (which requires a minimum sample size of 128 across the two groups). It should be noted that for the sample size calculation we ignored the increase in power that would occur due to the inclusion of the covariates in the analysis (Miles, 2003). We did this to ensure that the power analysis erred on the side of conservatism, as the effect of the covariates was unknown at this stage.

Testing

All pupils were tested at three points: before the intervention began (pre-test), between the completion of the program by the intervention group and entry of the controls into the program (first post-test), and again after all pupils had received the program (second post-test). The purposes of the three phases of testing were:

- pre-test – to obtain pre-test scores so the scores would be used as a covariate in the statistical analysis of post-test results;
- first post-test – to detect any difference in gains between experimentals and controls;
- second post-test – to detect whether the first cohort maintained any gains or even continued to make progress, or whether there was any tendency for their gains to

wash out; also to see whether the second cohort made gains similar to those of the experimentals (although it should be noted that this is not a randomised comparison).

Implementation of intervention

Each pupil received the program for one hour a day for 10 consecutive school days. The program was delivered by staff of the school trained by the designer of the computer program. To avoid the possibility that control pupils might receive the intervention early (i.e. contamination) all pupils received it via laptop computer. Because only enough laptops were available to deliver the program to six pupils at a time (30 per day), each cohort was randomly divided into three batches; the first batch of the intervention group received the program in the first fortnight and their controls received it in the fourth fortnight and so on. Pupils received the intervention *in addition to* their normal teaching; therefore, whilst sometimes pupils received the program during normal English classes, they also gave up time from other lessons to undertake the intervention. Overall, while receiving the program pupils received slightly less normal classroom teaching in English, but substantially more teaching in English than usual because of receiving the program.

All participants undertook the pre-test before any pupils received the program, and the second post-test after all the pupils had received it. However, the first post-test was given to each batch of the intervention group and their controls immediately after the intervention group finished the program, thus making three occasions of administration for the first post-test. The purpose of this was to ensure that all of the pupils in the intervention group received the first post-test before any wash-out could occur; each experimental batch's controls had to receive the first post-test at the same time as their parallel group.

Because access to the laptops and the program was strictly controlled and supervised, leakage between the cohorts was minimal. Also, the brevity of the program and the short timescale of the project meant that the pupils in the second cohort would not suffer any significant educational disadvantage from the delay in receiving the program (if demonstrated to be effective).

Instruments and data analysis

Both reading and spelling were assessed by members of the research team using age-appropriate standardised tests not previously used in the school. All tests were administered by researchers blind to individual pupils' group membership.
The tests used were:

- *Group Reading Test 6–14* (published by nferNelson), forms X (at pre-test and second post-test) and Y (at first post-test). This is standardised for ages 6–14, takes approximately 30 minutes, is group-administered, and can be given at any time of year;

- *British Spelling Test Series* (published by nferNelson), Level 3, forms X (at pre-test and second post-test) and Y (at first post-test). This is standardised for ages 9–15, takes 30–40 minutes, is group-administered, and is best given at the beginning of the school year, as we did.

The tests were, to the best of our judgment, fully appropriate measures of relevant literacy skills for pupils of this age, and therefore valid measures of outcomes which the intervention could reasonably be expected to effect.

At all stages the tests were marked by persons blind to individual pupils' group membership and independent of the school and the makers of the program. Similarly, the results were analysed by a statistician who was blind to individual pupils' group membership and was independent of the school, the software, and the University of Sheffield. To maintain confidentiality, data sent to the University of York did not contain any identifying details of individual children. Consistency of marking was ensured by double-marking of all scripts.

Statistical analysis

Our principal outcome was between-group differences in spelling scores at first post-test after adjusting for baseline spelling score, age and gender. We adjusted for pre-test means using analysis of covariance. For this outcome $p = 0.05$ was deemed to be statistically significant. The secondary outcome was differences in reading scores at first post-test. Again for this analysis the pre-test score was used as the main covariate but we also adjusted for age and gender. The p value deemed significant for this comparison was $p = 0.025$ to allow for multiple testing (i.e. to reduce the possibility of a Type I error, we need to reduce the value of the p value when undertaking more than one statistical test). We also followed up the children with a second post-test after the control group had received the intervention.

We also undertook an analysis to assess whether there was any effect of clustering on the outcomes. Because the children in the control group were being taught spelling and reading in their usual classes this might have led to statistical clustering of their outcomes by class. We undertook an analysis that took this into account by incorporating school class as a factor in the analysis. However, because the results were essentially the same as in the simpler linear approach, only the latter results are presented. For some children we had post-test scores but no pre-test scores. To assess whether this made any difference to the results we checked the main effect size using regression analysis but only adjusting for age and gender. All mean differences are presented along with their associated 95% confidence intervals.

Results

A total of 155 pupils were originally randomised to the study. Four children did not start the school and were not included in the study. Ten were not available for pre-testing and 17 did not attend for (first) post-testing. Figure 1 shows the distribution

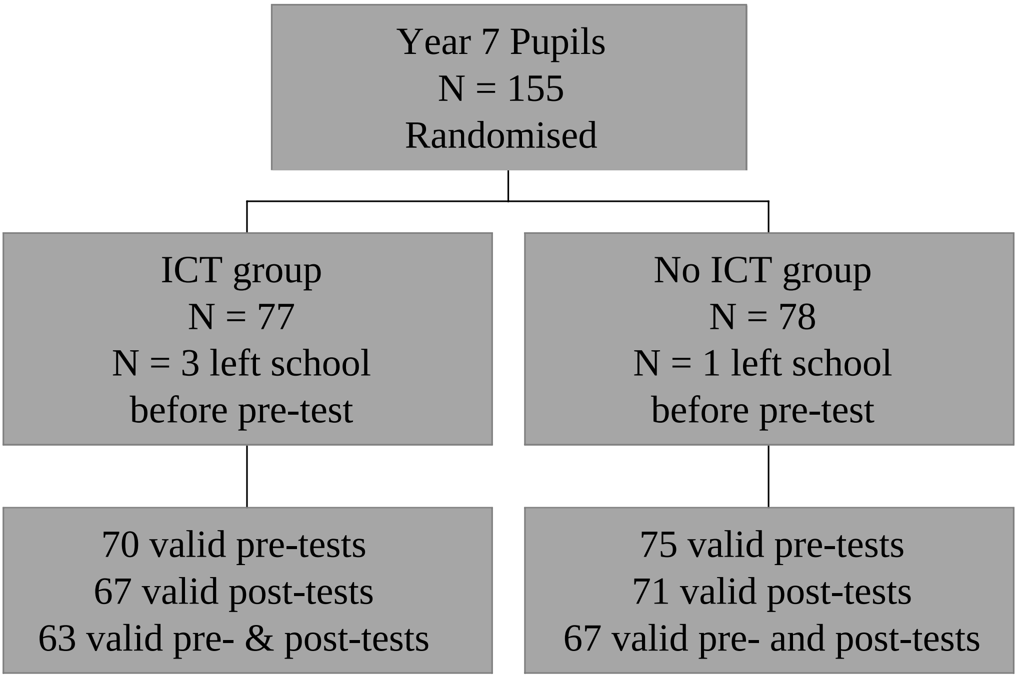

Figure 1. Participant flow in the ICT trial.

of these drop-outs. As the figure shows, there appear to be similar drop-outs between the groups. For the children who were missing for post-test their mean pre-test scores were not significantly different from the group that provided complete data (for reading: dropout mean 21.9 (sd 6.8), non dropout mean 18.3 (sd 8.2) p = 0.10; for spelling, dropout mean 28.5 (sd 16.7), non-dropout mean = 36.7 (sd 18.5), p = 0.08).

Table 1 describes the baseline characteristics of the two groups. As the table shows, there was some slight chance imbalance at baseline in pre-test spelling and reading scores, which we adjusted in the main analysis using analysis of covariance. The

Table 1. Baseline characteristics of randomised groups

	ICT Group (n =70)	No ICT Group (n =75)
Mean (SD) Age at 01/09/04	11.58 (0.30)	11.59 (0.31)
Number (Percent) girls.	37 (52.9%)	29 (38.7%)
Mean (SD) Pre test spelling raw score	38.47 (18.22)	33.33 (18.45)
Mean (SD) Pre test spelling standardised score	103.52 (14.85)	99.85 (13.54)
Mean (SD) Pre test reading raw score	22.26 (6.58)	20.68 (7.55)
Mean (SD) Pre test reading age	11.01 (1.57)	10.82 (1.56)
Mean (SD) Pre test reading standardised score	98.71 (11.35)	95.53 (12.85)

Table 2. Mean spelling and reading scores at first and second post-tests

	Variable	ICT Group Means (SD)	No ICT Group Means (SD)	Adjusted Difference* (Intervention – Control) (95% CI)	P value of adjusted Difference
1st Post-test	Spelling	38.57 (18.00)	37.24 (18.33)	0.954 (−1.83, 3.74)	0.500
	Reading	18.97 (6.11)	20.38 (6.64)	−2.33 (−0.96, −3.71)	0.001
2nd Post-test	Spelling	39.19 (15.24)	38.85 (18.37)	−2.54 (−8.13, 3.05)	0.370
	Reading	23.10 (8.03)	23.60 (7.84)	−1.32 (−3.95, 1.31)	0.321

*Differences are adjusted for pre-test spelling or reading score, age and gender.

N.B. By the 2nd post-test the 'No ICT' group had received the intervention.

children had a mean age of 11.5 years with an average reading age slightly below their chronological age.

Table 2 shows the means and differences between the groups at post-tests and the 95% confidence intervals. Both groups showed a small improvement in spelling but a slight decrease in reading scores. As the table shows, the control group tended to have slightly lower scores in spelling at the first post-test compared with the intervention group, but this difference was not statistically significant. In contrast, for our secondary outcome – reading scores – there was a statistically significant difference, with the ICT group having lower reading scores than the no-ICT group.

Some children were not present for the pre-test but we had post-test scores (n = 8). In a sensitivity analysis we undertook an analysis of covariance using age and gender only as the covariates. The observed mean differences at the first post-test were not materially different from the analysis that included just the children with both pre- and post-test data (Table 2).

Both groups were followed up again after the control group had received the intervention. Both groups showed an improvement in reading and spelling, but as table 2 shows, after adjusting for pre-test scores the slight differences between the groups were not statistically significant (although note that this was not a randomised comparison).

Discussion

We report, to our knowledge, the largest UK-based randomised controlled trial of using ICT for the teaching of spelling and reading among students of school age. We did not find a significant benefit of using ICT either in spelling or reading outcomes. Indeed, we unexpectedly found that those children receiving the ICT intervention had lower reading outcomes than the control group. Nevertheless both groups of children improved in spelling outcomes compared with their pre-test scores.

Computer programs are widely used both in schools and at home with a view to helping children learn to spell and increase other literacy outcomes. As noted previously, few recent rigorous evaluations have been undertaken. Ideally, all computer software ought to be evaluated using a randomised controlled trial design. This present study illustrates the importance of evaluating educational interventions using a randomised controlled design rather than the more widely used pre- and post-test method.

Our study has a number of weaknesses. First, the study is set in only one English comprehensive secondary school and it is possible that the results might have differed if the ICT intervention had been applied in a different setting. Secondly, we only evaluated one particular literacy software package designed to improve specific basic elements of reading and spelling and clearly our results are not necessarily applicable to other software packages, or this program when used with other age groups or when using different assessment measures. Thirdly, the possibility that our results were due to the children being withdrawn from normal school lessons cannot be excluded. Finally, although our sample size was relatively large, we cannot exclude the possibility of chance or some other unknown factor explaining our results.

On the other hand, our study has a number of important strengths. First, and most importantly, we used a randomised controlled trial design, which controls for selection bias and temporal changes. Second, it was set within a usual school setting and used in a way that followed normal school practice. Third, we re-tested the children soon after they had completed the ICT training and when we would expect, therefore, any intervention effects to be at their peak. Fourth, we avoided 'contamination' or leakage between pupils by restricting the availability of the software to laptop computers that were only available to pupils as part of the trial. Fifth, our trial was the largest to date and had sufficient statistical power to observe moderate benefits – if they had been present.

There is little robust evidence that demonstrates that ICT has a benefit on non-ICT learning in a school setting. The British Educational Communications and Technology Agency (BECTa) reports have used non-randomised data and shown little evidence for benefit (BECTa, 1998; BECTa, 2002). Similarly, a large non-randomised controlled trial in Israel did not detect an improvement in Hebrew literacy in the intervention schools but detected a significant reduction in maths scores compared with control schools (Angrist & Lavy, 2002).

This study supports the findings of the Torgerson and Elbourne (2002) systematic review, which did not find a statistically significant difference in effectiveness between the use of ICT and traditional paper and pencil methods for improving children's spelling abilities.

In conclusion, this randomised controlled trial did not detect any benefit of ICT for the learning of spelling or increasing reading skills using the test material adopted for the study. Before schools adopt ICT packages for teaching spelling or other literacy skills, these need to be evaluated in randomised controlled trials.

Acknowledgements

We thank Professor Sir Iain Chalmers for his helpful comments on an earlier draft of the paper. The authors wish to thank the author of the package we evaluated for commissioning the evaluation, Dr Maxine Burton of the University of Sheffield for administering many of the tests, and Jacquie Gillott and Colleen Woodward of the University of Sheffield for test marking and data entry. The authors all gave their time free. The University of Sheffield met the direct costs.

References

Angrist, J. & Lavy, V. (2002) New evidence on classroom computers and pupil learning, *The Economic Journal,* 12(October), 735–765.

Berninger, V., Abbot, R., Rogan, L., Reed, E., Abbott, S., Brooks, A., Vaughan, K. & Graham, S. (1998) Teaching spelling to children with specific learning disabilities: The mind's ear and eye beat the computer or pencil, *Learning Disability Quarterly,* 21, 106–122.

British Educational Communications and Technology Agency (BECTa) (1998) *The UK Evaluations Final Report* (Coventry, UK, BECTa).

British Educational Communications and Technology Agency (BECTa) (2002) *The Impact of Information and Communications Technologies on Pupil Learning and Attainment* (Coventry, UK, BECTa).

Brooks, G. (2002) *What Works for Children with Literacy Difficulties? The Effectiveness of Intervention Schemes* (London, D*f*ES Research Report no. RR380).

Heise, B.L., Papalewis, R. & Tanner, D.E. (1991) Building base vocabulary with computer-assisted instruction, *Teacher Education Quarterly,* 18, 55–63.

Lin, A., Podell, D.M. & Rein, N. (1991) The effects of CAI on word recognition in mildly mentally handicapped and non-handicapped learners, *Journal of Special Education Technology,* 11, 16–25.

Lipsey, M.W. & Wilson, D.B. (1993) The efficacy of psychological, educational, and behavioral treatment: confirmation from meta-analysis, *American Psychologist,* 48(December), 1181–1209.

McClurg, P. A. & Kasakow, N. (1989) Word processors, spelling checkers, and drill and practice programmes: effective tools for spelling instruction? *Journal of Educational Computing Research,* 5, 187–198.

Miles, J. N. V. (2003) A framework for power analysis using a structural equation modelling procedure, *BMC Medical research methodology,* 3(27), 1–11.

Torgerson, C.J. & Elbourne, D. (2002) A systematic review and meta–analysis of the effectiveness of information and communication technology (ICT) on the teaching of spelling, *Journal of Research in Reading,* 25, 129–143.

Torgerson, C.J. & Torgerson, D.J. (2001) The need for randomised controlled trials in educational research, *British Journal of Educational Studies,* 49, 316–328.

Torgerson, C.J. & Zhu, D. (2003) A systematic review and meta-analysis of the effectiveness of ICT on literacy learning in English, 5–16, in: *Research Evidence in Education Library* (London, EPPI-Centre, Social Science Research Unit, Institute of Education, University of London).