1

The impact of high access to computers on learning in mathematics, English and science

J. Gardner, H. Morrison and R. Jarman

School of Education, Queen's University of Belfast

Abstract   Researchers at Queen's University, Belfast recently completed a study into the potential of portable 'laptop' or 'notebook' computers in schools. Over two hundred and thirty pupils in nine schools were provided with a personal portable computer for a whole school year. One aspect of the research was to assess the impact which the high access to information technology (IT) had on the pupils' learning. Five experimental/control class groups (with/without laptops) were matched for age, gender and ability. The performance of these pupils in mathematics, English and science tests was measured before and after the 'treatment' period and the comparisons were analysed. A number of interesting effects were observed and these indicated, with due recognition of the project constraints, that the impact of high access to computers on learning in mathematics, English and science was at best marginal.

Keywords: Portable computers, mathematics, English, science, ANCOVA

Introduction

Sponsored by the Department of Education for Northern Ireland, the research set out to evaluate a number of issues relating to the rôle and impact of information technology, and in particular the up-and-coming generations of portable computers, in education. The project was broad ranging in terms of its context and research objectives.

The guiding principle of the evaluation design was to ensure that the methods would be 'fit for purpose'. In order to gain first-hand knowledge of classroom processes involving laptop usage, observations of lessons were carried out by the research team while a more in-depth and realistic teacher perspective of classroom processes was sought using techniques drawn from 'action research', in particular the use of diaries.

Conventional inquiry techniques such as interviews and questionnaires were used to develop and refine the perspectives of the teachers, school principals and a selection of parents of the pupils with laptops. Pupil diaries also provided information on the pupils' use of their machines in school and at home. In support of these qualitative methods, a quasi-experimental design, involving the testing of matched pairs of experimental (with laptops) and control (without laptops) groups, was used to assess the impact of the use of the machines on pupil performance in English, mathematics and science. Attitudes to these disciplines and to school in general were also assessed using questionnaires.

The analysis of these various types of data provided the basis for a professional appraisal of the 'worthwhileness' of having personal, portable computers. For full details of the project's general findings in relation to curriculum fit, classroom processes, implications for teaching, the views of management and parents, pupil attitude analysis, hardware and software issues and the research design and methods please refer to Gardner et al 1993. This paper focuses on the results of the research in relation to the impact on the pupils' academic performance.

Scope of Project

The project involved nine schools: one special education school, one primary school, six non-selective secondary schools and one selective (grammar) school[1]. In each school one whole class was supplied with laptops with, in most cases, at least two extra machines for the teachers. While the pupils and their various teachers were encouraged to use the machines throughout their curriculum, their participation was specifically monitored in a 'focus' subject: English, mathematics or science. The teacher of this focus subject (referred to as the 'focus teacher') acted as the person liaising with the project team and as the teacher researcher for the classroom and curriculum-based aspects of the research. The schools and their focus teachers were made ready for the project during the period April 1 to June 30, 1991. This involved in-service training (INSET) courses for the focus teachers and meetings with the parents of the pupils who were to be involved in the project. The fieldwork for the project began in September 1991 and was completed in late June 1992.

There were 235 pupils in the experimental groups (those with laptops) and 191 pupils in the control groups (those without laptops). The details of the laptop classes, including the machine types used, are listed in Table 1 below:

Table 1: Details of Participating Schools and Classes

School	Type	Gender Mix	Age Range	Year[2] Group	Key[2] Stage	Focus Subject	No. of Pupils	Machine[3]
1	Non-Sel	Co-Ed	13-14	10(9)	3	Mathematics	19	NB 201
2	Special	Co-Ed	13-14	10(9)	3	English	6	T1000SE
3	Non-Sel	Girls	14-15	11(10)	4	Mathematics	26	T1000SE
4	Non-Sel	Girls	12-13	9(8)	3	Science	28	T1000SE
5	Non-Sel	Boys	13-14	10(9)	3	Mathematics	24	NB 201
6	Non-Sel	Co-Ed	13-14	10(9)	3	Mathematics	29	Powerbook 100
7	Non-Sel	Boys	12-13	9(8)	3	English	28	T1000SE
8	Primary	Co-Ed	10-11	7(6)	2	Primary	24	NB 201
9	Sel	Co-Ed	12-13	9(8)	3	Science	<u>26</u>	NB 201
							210	

(Note: 'Non-Sel' = non selective, 'Sel' = selective, 'Co-Ed' = co-educational
'Year Group' is the N.Ireland designation, England and Wales designations are in parentheses)

A small number of machines (1-2) were also provided for the teachers in each school while a further twenty-five machines, mainly Tandy WP2s, were distributed to three small groups of remedial pupils for whom no control groups existed.

The project demonstrated that the NB201, TS1000SE and Powerbook machines were functionally capable of providing for all the IT-related aspects of the English and mathematics curricula and the data capture aspects of the science curriculum. The main software used was the integrated wordprocessor, database and spreadsheet package (Microsoft Works on the IBM compatibles with Claris Works on the Apple Powerbooks), a Logo-like package, Qlogo, commissioned by the researchers for both platforms and the 'Sense and Control' (Educational Electronics Ltd) package for science. The pupils reported using their machines in a wide range of subjects including: French, German, Latin (one school), computer studies, information technology, geography, business studies, commerce, history, religious education and careers. The vast majority of this work involved

wordprocessing but work in some subjects such as geography, history and computer studies involved the use of spreadsheets and databases.

Cross-curricular usage, in the sense of theme work spread over two or more subject areas, was limited but instances were recorded of the gathering of data in history, home economics and science classes for use in mathematics lessons. In another instance desktop publishing of work in history lessons was facilitated in English.

Activities with databases or spreadsheets, and to a lesser extent QLogo were essentially teacher-directed, often with verbal or worksheet-based step-by-step instructions. This would appear to be necessary as the multi-function/multi-command nature of these systems are less easily assimilated and remembered by the pupils than the processes involved in wordprocessing. Interesting examples of the pupils using the laptops independently and purposefully were recorded in science and mathematics. For the most part in most classes, pupil-inspired work tended to involve wordprocessing with, to a lesser extent, exploratory work using the QLogo software. In most cases the classroom usage decreased significantly as seasonal and end-of-term activities (eg. school trips, plays and carol services) took their toll of normal work. By the end of the project the variation in pupils' personal use of the machines ranged from 1 to 20 hours per week but it should be noted that the upper range was significantly influenced by the incidence of 'games' playing.

Experimental Design and Procedures

The cognitive research hypotheses proposed that post-test scores of the experimental (laptops) group in each of the three disciplines (mathematics, science and English) would be superior at the 5% significance level to those of the control group when the scores were adjusted for differences in pre-test scores.

The research design required that the schools provide two classes which were following the same courses in the same year group. According to the school type each group was categorized as being from a selective or non-selective school and as being single-sex or co-educational. Although desirable for proper matching it was not possible to randomly assign the pupils to the experimental and control groups as the classes had been fixed by the schools. The research design should therefore be accepted as quasi-experimental.

Matched class pairs

The schools had been encouraged to offer groups which, in their estimation, were matched for ability. To increase the reliability of the comparison of 'like versus like', the pupils' reasoning ability scores were measured at the start of the project using established tests from the National Foundation for Educational Research (NFER) 'AH' series. The analysis produced four secondary experimental/control group pairs which were matched for age and gender and which did not differ significantly in reasoning ability scores. The control and experimental classes in three of the secondary-level schools were found to be matched while the fourth pair were taken from two separate schools. In order to provide as much information as possible the results for the primary school matched pair are also presented but since these were matched at the less secure level of p=0.0702, they should be treated with some caution. The matched pairs formed the basis of the analysis of performance in mathematics and science.

The matched pairs (mpA through to mpE) are summarized thus:

Table 2: Details of Matched Pairs

Pair	School Type (Gender)	(Selection)	Age Range (Key Stage)	No. of Pupils	Mean AH Scores	F	p
mpA	Co-Ed	Sel	3	26,24	73.15, 75.33	0.5851	0.4480
mpB	Girls	Non-Sel	3	28,30	63.54, 63.90	0.0186	0.8921
mpC	Co-Ed	Non-Sel	3	29,31	64.52, 68.59	2.1474	0.1486
mpD *and*	Boys	2 x Non-Sel	3	28,21	52.15, 48.05	1.2258	0.2742
mpE	Co-Ed	Primary	2	24,24	34.83, 31.12	3.4320	0.0702

('No. of Pupils' and the 'Mean AH Scores' are presented in the form: experimental, control
'Non-Sel' = non selective, 'Sel' = selective, 'Co-Ed' = co-educational)

For the English analysis, a 25% proportionate random sub-sample of all of the secondary-level school pupils, together with all of the pupils in both primary classes, were selected for the experimental and control groups. There was no significant difference in reasoning ability for either the secondary-level sample (F = 0.3198, p = 0.5746) or the primary sample (F = 1.9115, p = 0.0702).

Performance measures

The performance of the matched pairs was examined using pre- and post-tests in mathematics, science and English with the tests in the latter taking the form of narrative writing tasks. The pre- and post-tests were administered in the first six weeks and the last eight weeks of the project respectively. The same test (ie. the 'test/retest' method) was used for mathematics and science while 'parallel' tasks were used for English. The gap between the pre- and post-test administrations was approximately eight months in each school.

Method of analysis

Some explanation of the techniques used to analyse the results is necessary. Aside from the chosen method, analysis of covariance - ANCOVA, two other methods suggest themselves for comparing the performance of the experimental and control groups. The first of these is 'gains analysis'. It was considered inappropriate to merely conduct an analysis of variance (ANOVA) of the performance gains of the experimental group against those of the control group. Average performance gains so measured suffer from bias towards the lower performing pupils (both experimental and control) since those who score low in the pre-test have greater scope to improve in the post-test. Similarly it is not tenable that a gain from 20 to 25 for one pupil should be be considered equal to a gain from 80 to 85 for another. The use of derived gain score data, such as residualized gain scores which do not offer true comparisons, were also deemed inappropriate.

The second method is to convert the pre- and post-scores to within-subject (ie. for each pupil) independent variables. The difficulty with this approach is that the impact of having personal access to a laptop computer would be measured as a secondary interaction rather than as the main effect. Both methods were rejected in favour of the greater rigour afforded by ANCOVA methods.

ANCOVA

The conventional method for ensuring maximum reliability for comparison of pre- and post-test scores is to use analysis of covariance (ANCOVA) to adjust the post-scores for any differences between the pupils in the pre-tests. ANCOVA is an extension of the technique of the analysis of variance in which the effects of the independent variable (being in a laptop or a control class) are measured after the dependent variable (the post-test scores in mathematics, English or science) is adjusted for differences associated with the covariate (the pre-test score in mathematics, English or science). In other words, ANCOVA addresses the following question:

> Is there a significant difference (other than that which might be expected by chance) between the laptop and control classes, as measured by the post-test score (in any of the three core disciplines) after this score is adapted for differences in pre-test scores?

Analysis of covariance is particularly appropriate in quasi-experimental designs where subjects cannot be randomly assigned to treatments. ANCOVA adjusts the group means to what they would be if all subjects scored identically on the covariate (pre-score). It is then possible to attribute differences between subjects to the effects of the independent variable (here personal access to a portable computer) and not to differences between the subjects on the covariate (pre-score). All analyses were carried out using the statistical package, SPSS.

Reliability

Reliability of the covariate measure is vital to ANCOVA as poor reliability will result in a loss of power and under-adjustment of the error term. In non-experimental research poor reliability can lead to type I and type II errors. In this work high reliability (ie. $r_{xx} > 0.8$) was guaranteed for mathematics and science as detailed below but no similar assurance can be offered in the case of English as no calculations based on generalizability theory (Godshalk et al 1966) have been carried out. The tables and graphs below (from Table 3 onwards) present descriptive statistics in terms of gains but it should be stressed that ANCOVA was the means of analysis and not analysis of variance (ANOVA) of gain scores.

Test administration

All test administration and marking was carried out by the research team with the exception of the measure of ability in English. This was marked by the research team but was administered by the teachers.

Measure of general reasoning ability

The non-verbal reasoning ability of the primary school pupils was assessed using the NFER AH1 test; a test which measures skills in dealing with series, likes, analogies and choices using pictorial and diagrammatic items in a multiple choice format. The general reasoning ability of the secondary and grammar pupils was measured using the 42 minute version of the NFER AH2 test. In addition to a total measure of general reasoning, the AH2 test offers a profile in the three reasoning areas: numerical, verbal and perceptual.

Measure of ability in mathematics

The instruments used to measure the mathematical ability of pupils were constructed by the University of London IMPACT research team (IMPACT 1992) from a bank of University of Wisconsin 'superitems'. A superitem is an item which, on the basis of one stem problem, asks questions at four levels of difficulty. The assessment of both primary and secondary pupils was based, in the main, upon the same items except that primary pupils were not presented with the extended level (level four) of difficulty. The primary test consisted of 24 items while the secondary test was a 37 item test. The reliability of both tests was measured for the pupils in the sample using Cronbach's alpha coefficient (Cronbach 1951,1970). In both cases alpha exceeded 0.87, indicating high internal consistency for the tests. It is worth noting however that as some of test items did not comply with the facility and item-total correlation ranges considered appropriate for such tests, the test had to be viewed as measuring a global construct ie. general mathematical ability; and could not be broken down to look at ability in specific topics such as Algebra and so on.

Measure of ability in science

Ability in science was measured using items drawn from the Assessment of Performance Unit's science surveys conducted in the period 1980 to 1984 (APU 1980). All items were chosen from *Category 1: Use of Graphical and Symbolic Representation*. The primary science test comprised 16 items selected from the Age 11 and Age 13 banks and the secondary test comprised 17 items chosen from the Age 11, Age 13 and Age 15 banks. As was the case with mathematics, Cronbach's alpha was computed for the full primary and secondary samples and in both cases alpha exceeded 0.96, indicating very high internal consistency.

Measure of ability in English

Ability in English was measured by providing pupils with the opening sentences of a story and asking them to complete the story. Pupils were permitted to write for up to 50 minutes and the quality of writing was then assessed using both atomistic and holistic measures. Six atomistic measures were used to assess each essay:

- total number of words;
- number of full-stops omitted per 100 words;
- number of spelling errors per 100 words;
- number of possessive apostrophes used correctly per 100 words;
- number of incorrect apostrophes per 100 words
- number of correct 'missing letter' apostrophes (eg. between the *n* and *t* of *doesn't*) per 100 words

This data was augmented by holistic scores including:

- an impression mark, scored in the range 1 to 7;
- Assessment of Performance Unit's measures, in a range 1 to 5, of content and organization, appropriateness and style, grammatical conventions and orthographic quality; and
- modified Torrance measures of creativity, in a range 1 to 5, of *fluency* - no of ideas, characters etc, *flexibility* - variety of vocabulary, sentence beginning etc, *originality* - new ideas, insights, strange or surprising relationships and *elaboration* - detail, made fancy etc.

Given the creative nature of English language, a measure of the reliability of the English 'test' would be problematic but measures were taken to maximize the reliability of the marking. All of the scripts were marked by one person and 10% were marked by three other markers for triangulation purposes. The criteria for APU and modified Torrance creativity marking were established by a group discussion of the marking team. All the markers are

currently or have been practising English teachers. The APU marking (scoring 1 to 5) was impressionistic while the Torrance marking comprised two marks for each of two criteria plus one mark for overall impression. The markers were not aware which scripts were from experimental pupils and which scripts represented the work of control pupils.

Problematics

The overriding constraint for the experimental aspects of the research was the relatively short timespan of one school year in which learning enhancement was to be measured. The period was one school year (comprising approximately 185 working days, nett of holidays) and was shortened at both ends as result of the 'settling in' period at the beginning and the 'winding down' period leading to the summer holidays at the end. In operational terms all of the schools experienced a normal school year but it should be remembered that the work of the project was set against a backcloth of major curricular change in all schools.

Regression analysis revealed that, after pre-score, reasoning ability was the main predictor of performance in the tests and was therefore controlled throughout. Teaching quality, the next most important likely influence on performance, can rarely be controlled for and this project was no exception. The influence of teacher characteristics such as the level of their commitment to fulfilling the IT-related curriculum requirements, their teaching ability (in terms of such things as their power to motivate pupils, personal enthusiasm for their subject, their ability to innovate, their classroom management skills and so on) and their commitment to the project is open to debate.

It is also possible that the performance of one of the matched pairs (mpA - selective) could have been affected by machine difficulties. The school experienced problems with the machine they had been supplied with to the extent that work with the laptops was curtailed to a 'ticking over' level for most of the months of December 1991 and January 1992 until the supplier replaced them. The other pairs experienced minimal difficulties by comparison. Examination of the results below actually shows that the experimental group in mpA performed better than the other experimental groups so any likely effect here is difficult to identify.

Some school effects may also be considered as possible influences on the results. These include the Hawthorne effects which attached to the experimental groups. Right from the start the laptops groups felt themselves to be special in relation to their peers and this perception was shared by other pupils and indeed some teachers. It is possible that this effect might have raised the expectations and performances of the pupils and may have affected the way in which the teachers worked with the experimental classes. It is less likely that such effects could diminish the pupils' performances.

As will be argued later, the acceptance of the null hypothesis suggests that all of these effects, with the exception of the timespan, can be discounted as significant influences on the results.

Summary and Discussion of Results

The results for the measures of impact on performance in the mathematics, science and English tests are presented in the following tables and summary plots. (The reader may wish

to refer to Table 2 for details of the matched pairs). The pre-scores, post-scores and mean gains are provided for the experimental and control groups of each matched pair and the significance of any difference between them is noted. The descriptive statistics, including gains, are offered as illustration of the comparisons made. *It should be emphasized that the analytic technique was ANCOVA, with pre-score as the covariate, and not an ANOVA of these gain scores.*

Table 3: Results for the Matched Pairs in Mathematics and Science

		Mathematics					Science				
Matched Pair		Exp'tal Means	SD	Control Means	SD	Sig	Exp'tal Means	SD	Control Means	SD	Sig
A	Pre	20.81	3.42	24.37	4.25		79.90	6.16	84.21	3.88	
	Post	24.04	4.54	25.42	4.13		85.24	7.52	84.25	7.32	
	Gain	**3.23**		**1.05**		NS	**5.34**		**0.04**		p<0.05
B	Pre	18.89	3.70	17.89	4.83		67.52	9.37	70.83	9.38	
	Post	19.63	3.73	19.89	4.12		75.70	10.10	74.38	10.06	
	Gain	**0.74**		**2.00**		NS	**8.18**		**3.55**		NS
C	Pre	23.37	3.68	23.41	3.35		81.23	6.95	82.75	6.79	
	Post	25.22	3.85	26.28	3.23		86.50	5.68	86.44	3.78	
	Gain	**1.85**		**2.87**		NS	**5.27**		**3.69**		NS
D	Pre	14.92	3.38	12.89	4.39		54.61	10.58	51.13	10.35	
	Post	16.19	4.34	16.39	5.34		62.71	12.15	53.87	13.90	
	Gain	**1.27**		**3.50**		NS	**8.10**		**2.74**		NS
E	Pre	13.00	6.67	6.96	3.38		48.45	21.59	32.26	13.57	
	Post	15.04	5.89	9.12	3.26		55.86	20.18	38.58	15.16	
	Gain	**2.04**		**2.16**		NS	**7.41**		**6.32**		NS

(SD = Standard Deviation, Sig = Significance, NS = non significant)

Mathematics

The results indicate a slight non-significant effect (illustrated in Figure 1) in favour of the control groups with the only exception being those in favour of the experimental group in the relatively high ability selective school matched pair (mpA). It is not unreasonable to conclude that in the circumstances of this project, the laptop usage did not impact favourably mathematical ability in general. This view has a number of implications which are worthy of consideration and further research. As the mathematics test instruments were tested and shown to be highly reliable measures of general mathematical ability the results raise the possibility that contrary to the received wisdom, high levels of IT usage may not impact on the pupils' abilities in mathematics, at least not in a one year timespan. It is still possible that the use of IT might have increased the pupils' performance in more focused IT-related aspects of mathematics (for example data handling) but the instruments used were not able to reliably test this.

***********Figure one here

Set against this theorizing are the observations that the teachers and pupils alike felt that they were behind, in comparison to other mathematics groups in the same year, in their mathematics teaching/learning schedules. A lag of up to three weeks was quoted by the teachers but it is questionable whether or not such a lag could adversely affect performance since it should only reduce the content covered rather than affect the skills and abilities developed. Perhaps more importantly was the perception of the IT and laptop activities as not being particularly mathematical. On a number of occasions pupils were asked about their spreadsheet work and they were not able to articulate reasons for using the tool nor did they show any understanding of the mathematical processes which underpin the formula facility for example. If anything they were more concerned with the software command sequences than the mathematical processes. In contrast pupils in English and science lessons were quite clear on the advantages afforded by wordprocessors and spreadsheet graph plotting respectively. There is some basis to the suspicion that the delivery of IT-related mathematics tends to be somewhat 'bolted on' ie. contrived outside the normal mathematical context and thereby distanced from the processes it is widely considered to enhance. Paradoxically, however, the attitude survey (not reported here) indicated that a greater proportion of the experimental pupils perceived mathematics as a relevant to the 'real world'. Perhaps the very use of the laptops in mathematics lessons might increase the perception of the relevance of mathematics itself through introducing a practical, 'worldly' element into the lessons.

Overall it has to be concluded that the 'treatment' ie. the personal access to the laptops, did not result in a significant difference between the performances of the experimental and control groups. The 'null hypothesis' is therefore upheld.

The comparison of the gains in the science results is illustrated in Figure 2:

***********Figure 2 about here

Science

The science test results indicated a positive non-significant effect in the performance of the experimental groups in comparison to their matched control groups. Indeed the results for the high ability selective school's matched pair (mpA) showed a statistically significant result (p<0.05) in favour of the experimental group. It is interesting to note that in contrast to the mathematics tests, the science test was much more focused on data handling (ie. IT-related) items. This leads to the suggestion that IT-usage does impact on this focused aspect of science education. The qualitative results suggest that the IT-related activities were perceived to be much less contrived as the pupils were engaged in 'real' science experimentation, with the computers being recognizably useful in logging data and processing it into graphs via spreadsheets. Wordprocessing the experiment write-ups may also contribute to a sense of purpose in the laptop usage which may not be as strong in mathematics contexts.

English

As might be expected the results, listed in Table 4, for the number of words written (remembering that the tasks were handwritten) show that the personal access to laptops had no significant impact on the amount of writing which the pupils are able to achieve in a given

timed task, however it is worthy of note that the primary children in the experimental group did show a much larger mean gain in the number of words they wrote. The teachers reported that wordprocessing had resulted in longer pieces of work than they would normally have expected from the pupils but that the increases were gained over a period of time involving redrafting and editing. Any direct comparisons of the amount handwritten vs the amount wordprocessed in any given timespan would be likely to be confounded by the pupils' acknowledged lack of typing skills.

Table 4: Mean Number of Words in the Writing Tasks

	Secondary					Primary				
	Exp'tal Group	SD	Control Group	SD	Sig	Exp'tal Group	SD	Control Group	SD	Sig
Prescore	552.26	215.22	489.17	161.75		323.94	119.84	224.25	71.05	
Postscore	532.48	133.59	513.26	165.26		486.35	146.97	263.60	143.80	
Gain	**-19.78**		**24.06**		NS	**162.41**		**39.35**		NS

(SD = Standard Deviation, Sig = Significance, NS = non significant)

The results for the various atomistic and holistic measures are presented below (Table 5) but note that three of the atomistic scores (number of possessive apostrophes used correctly; number of incorrect apostrophes and number of correct 'missing letter' apostrophes, eg. between the *n* and *t* in *doesn't* -all per 100 words) have not been included as their values were felt to be too small to have any reasonable meaning. Inspection of the means and corresponding standard deviations indicate the need for caution in interpreting the remaining atomistic scores. The reader should note that skewness and kurtosis measures of the atomistic scores, together with the small sample size of pupils, indicate the need for caution in interpreting the data in Table 5.

Table 5: Performance Measures for English

Measure	Secondary Pupils					Primary Pupils				
	Exp'tal Group	SD	Control Group	SD	Sig	Exp'tal Group	SD	Control Group	SD	Sig
Atomistic Measures										
A1: Full stops omitted										
Prescore	1.17	1.82	0.33	0.52		1.62	2.58	2.39	1.93	
Postscore	0.51	0.98	0.26	0.34		0.32	0.54	2.36	2.27	
Gain	**-0.66**		**-0.07**		NS	**-1.30**		**-0.03**		p<0.05
A2: Spelling errors										
Prescore	2.01	1.83	2.35	1.88		3.50	6.06	3.55	3.07	
Postscore	1.51	1.23	2.06	1.51		2.58	3.52	3.17	2.24	
Gain	**-0.5**		**-0.29**		NS	**-0.92**		**-0.38**		NS
Impression Marking I										
Prescore	4.78	1.04	4.30	1.40		4.53	1.66	3.00	0.86	
Postscore	5.65	1.23	5.43	1.16		5.06	1.95	3.10	0.97	
Gain	**0.87**		**1.13**		NS	**0.53**		**0.10**		NS
APU Assessments										
P1: Content and org.										
Prescore	3.91	1.04	3.65	0.98		3.82	1.47	2.80	0.89	
Postscore	4.35	1.07	4.26	0.81		3.82	1.59	2.45	0.94	
Gain	**0.44**		**0.61**		NS	**0.00**		**-0.35**		NS
P2: Approp'ness,style										
Prescore	3.87	1.18	3.39	1.23		3.76	1.44	2.75	0.91	
Postscore	4.22	1.13	4.17	0.94		4.06	1.25	2.35	0.99	
Gain	**0.35**		**0.78**		NS	**0.30**		**-0.40**		p<0.05
P3: Grammar										
Prescore	3.78	1.13	3.35	1.27		3.76	1.30	2.80	0.77	
Postscore	4.22	1.09	3.96	1.02		4.00	1.37	2.55	1.05	
Gain	**0.44**		**0.61**		NS	**0.24**		**-0.25**		p<0.05
P4: Orthographic										
Prescore	3.74	1.25	3.35	1.19		3.88	1.45	2.95	0.69	
Postscore	4.17	1.11	4.00	1.09		3.88	1.54	2.70	1.22	
Gain	**0.43**		**0.65**		NS	**0.00**		**-0.25**		NS
'Torrance' Measures										
T1: Fluency										
Prescore	4.13	0.97	3.52	1.34		4.00	1.27	3.30	1.13	
Postscore	4.65	0.71	4.39	0.78		4.06	1.30	2.80	1.24	
Gain	**0.52**		**0.87**		NS	**0.06**		**-0.05**		p<0.05
T2: Flexibility										
Prescore	3.87	1.07	3.35	1.37		3.41	1.37	2.00	0.97	
Postscore	4.35	1.03	4.17	0.89		3.94	1.43	2.70	1.13	
Gain	**0.48**		**0.82**		NS	**0.53**		**0.70**		NS
T3: Originality										
Prescore	3.52	1.31	3.13	1.32		3.47	1.42	2.10	1.21	
Postscore	4.30	0.93	4.13	1.06		3.76	1.64	2.25	1.25	
Gain	**0.78**		**1.00**		NS	**0.29**		**0.15**		p<0.05
T4: Elaboration										
Prescore	3.61	1.23	3.43	1.31		3.59	1.42	2.05	0.89	
Postscore	4.17	1.15	4.04	0.98		3.76	1.64	1.90	0.85	
Gain	**0.56**		**0.61**		NS	**0.17**		**-0.15**		p<0.05

(SD = Standard Deviation, Sig = Significance, NS = non significant, p = probability)

The results for the secondary sample English tasks (illustrated in Figure 3, for x-axis legend see Table 5) showed no statistically significant differences between the experimental and control groups but did present some interesting non-significant trends in the results. In the 'atomistic' measures (ie. number of full stops omitted per 100 words, number of spelling errors per 100 words) the experimental groups performed better while in the 'holistic' measures (including APU assessments and modified Torrance measures for creativity) the control groups performed better. Notwithstanding the difficulties with the atomistic measures it is reasonable to propose that the computer based drafting and redrafting activities and the usage of spell-checkers sufficiently increased the pupils' awareness of and practice in these skills to transfer to their handwritten work.

***********Figure 3 about here

Note: in interpreting Figures 3 and 4 (below), the A1 and A2 measures refer to spelling errors and 'omitted' full-stops so the negative gains indicate less of each, ie. an improved performance.

The lesser performance on the holistic measures would suggest, however, that the use of the computers did not significantly enhance the writing quality of the pupils and may, on the contrary, encourage a somewhat structuralist approach to the detriment of creativity and style. Virtually the opposite appears to be the case for the primary sample with many of the holistic measures showing significantly improved performance for the experimental pupils (illustrated in Figure 4)

***********Figure 4 about here

The English teachers were unequivocal in their view that the content and presentation of their pupils' work had been improved. They were also adamant that the quality and effort in the pupils work was exceptional in comparison to the expectations they would have had for the group under ordinary circumstances. In curriculum assessment terms, some of the pupils were considered to be working at a level two levels above that expected of them. Some of the primary pupils, for example, were adjudged to be attaining level 7 in the *Writing* attainment target (Northern Ireland Curriculum). Pupils were reported to be more prepared to experiment with their writing and be more confident about expressing themselves. In all cases the pupils were reported to write 'from the head' rather than prepare a handwritten piece for typing. Notwithstanding this many of the pupils reported in their diaries and in conversations with the researchers that they still preferred to handwrite for speed and convenience.

The overriding feature of the results is that there was no significant difference in gains, at the 5% level, in the large majority (19 out of 20 in fact) of the comparisons of the secondary experimental vs control data which were examined. The significant differences in the primary comparisons are worthy of note but it should be remembered that they refer to just one matched pair in tests in a subject, English, where the reliability of the measurements can only be optimized rather than guaranteed. With the lack of measured significant differences

the null hypothesis, that the treatment has had no significant effect on the experimental group's performance in mathematics and science, can be reasonably concluded to have been upheld overall.

The importance of the null hypothesis

Earlier it was claimed that the logic of the experimental design ensured that if the null hypothesis was upheld then the possible distorting effects (other than timespan of the project) were rendered irrelevant. To assist in explaining this the reader can consider the four possible options for the treatment effect and the influence of distorting effects upon it. These are: a sizeable positive effect, a marginal positive (or negative) effect, no effect and a sizeable negative effect. The last of these can obviously be dismissed on examination of the results.

Firstly for the null hypothesis to hold when there actually has been a sizeable positive effect then it must have been confounded and rendered not significant by some overriding negatively distorting effect. We have already said that once ability is controlled such an effect might be the teaching experienced by the pupils. The teaching of the experimental groups in comparison to that experienced by the control groups would have to be sufficiently poor to hinder and indeed reduce any gains. In this project the teaching quality for the experimental groups was not measured but was comprehensively observed. It is the view of the researchers that the teaching ranged from routine and adequate to highly accomplished and motivational, with the balance very much to the latter. The likelihood of a major negative teaching effect on the experimental groups can therefore be reasonably discounted.

Secondly, the null hypothesis does not rule out the possibility that there has actually been a positive or negative *marginal* effect from the treatment but it does imply that the effect was so small in terms of its overall impact on general mathematical ability, on data-handling in science and on the quality of writing in English that it was largely (ie. not in every case) indistinguishable from that which might be attributed to chance. Clearly if there is a marginal effect from the treatment then any positive or negative distorting effect could not be significant otherwise it would cause the groups to be significantly different and the null hypothesis to be rejected. In the case of such a marginal effect of the treatment, a longer experimental period might be needed to increase the effect sufficiently for reliable measurement.

Finally if there was actually no effect from the treatment, then discussions of distorting effects are not relevant.

Conclusions

The results are therefore considered sufficiently secure to conclude that for this project the impact of personal access to laptop computers on pupils' performance was not significant or, at best, was marginal over one school year. Further work would be required to investigate the influence which a longer time period might have.

Notes

1.   Northern Ireland, in common with several other parts of the UK, retains selective-entry grammar schools. Pupils are awarded places according to their performance in selection tests administered at 11 years of age.

2. The Northern Ireland Curriculum reform is similar in most respects to the National (England and Wales but not Scotland) Curriculum reform and the age range of pupils is similarly divided into year groups and 'key stages'. However as children in Northern Ireland begin primary education earlier than their counterparts in England and Wales, equivalent year group designations differ by one for the same age range. The 'key stages' are similar for both curricula and divide the primary and secondary education phases into age ranges: KS1 (approx 5-7/8 years), KS2 (approx 7/8-11 years), KS3 (approx 11/12-13 years) and KS4 (approx 13/14-16 years).

3. The machines used in the project were the Research Machines Ltd (a UK manufacturer) NB201s, Toshiba T1000SEs, Apple Powerbook 100s and Tandy WP2s. The first two of these are IBM-compatible with 1Mb RAM and 1.44 Mb 3.5 inch disk drive while the Apple Powerbook is compatible with desktop Macintosh machines. The Tandy WP2 is a relatively simple machine with proprietary wordprocessing software in ROM. The main software packages were supplied by Microsoft (UK) Ltd (Microsoft Works), Claris Corporation (Claris Works) and Educational Electronics Ltd (Bedfordshire, UK - 'Sense and Control')

References

APU (1980) Assessment of Performance Unit Science Item Banks, c/o National Foundation for Educational Research, The Mere, Upton Park, Slough, UK

Cronbach, L J (1951) Coefficient alpha and the internal structure of tests *Psychometrika* 16 297-334

Cronbach, L J (1970) *Essentials of Psychological Testing* New York, Harper and Row

Godshalk, F, Swineford, F and Coffman W E (1966) The measurement of writing ability. *College Board research Monographs* No. 6

IMPACT (1992) Research Project: 'The impact of information technology on children's achievements in primary and secondary schools', c/o Dr M Cox/IMPACT Team, Centre for Educational Studies, King's College London

NFER National Foundation for Educational Research, The Mere, Upton Park, Slough, UK

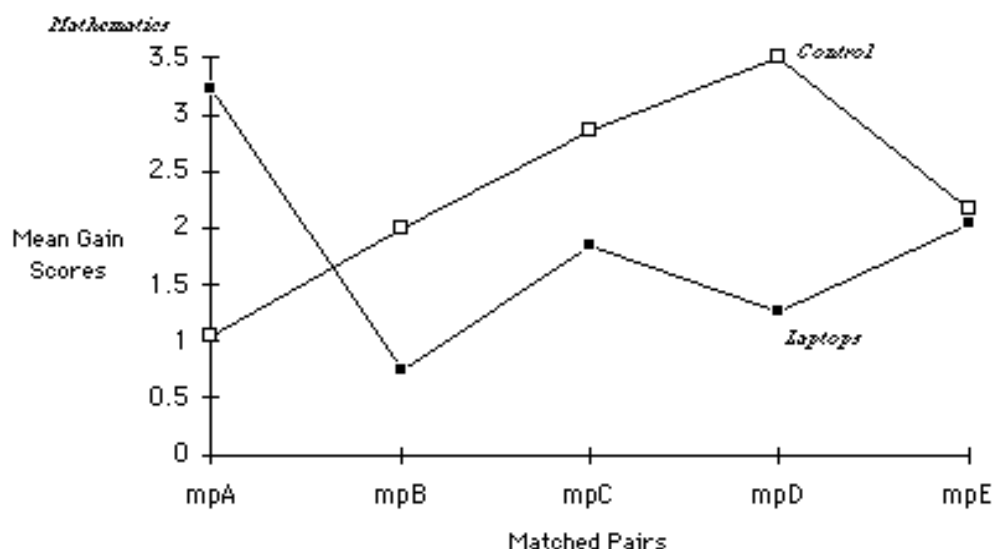Torrance, E P (1962) Guiding Creative Talent New Jersey, Prentice Hall

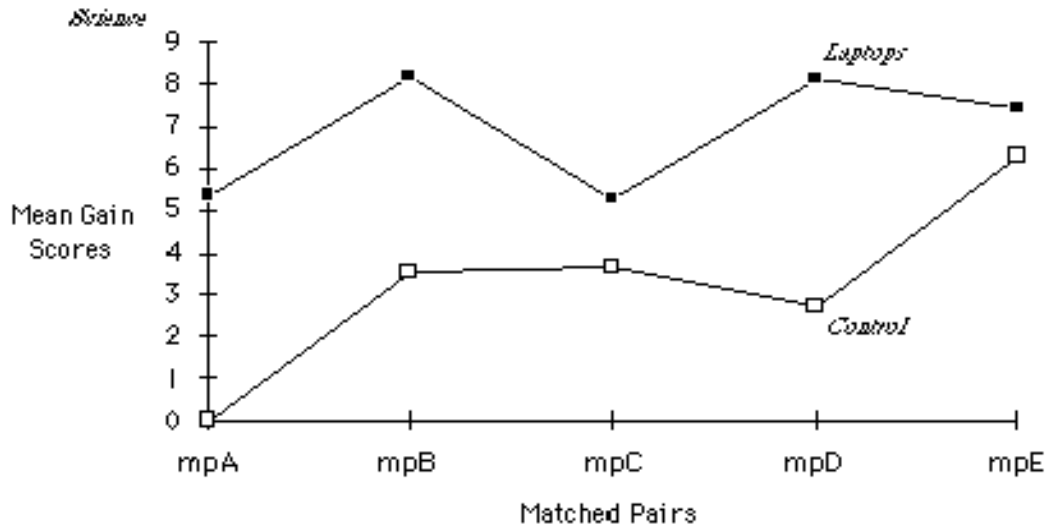**Figure 1:   Gain Scores for the Matched Pairs in Mathematics**

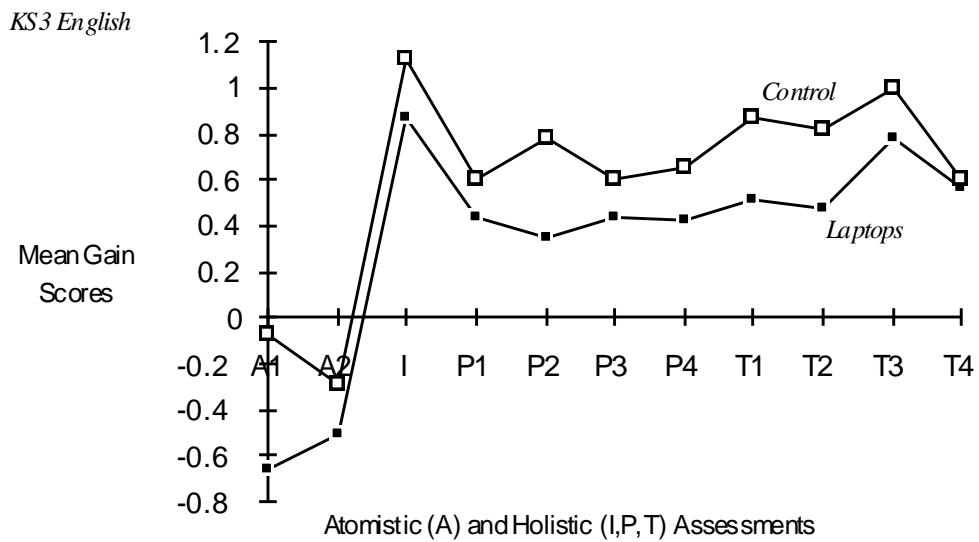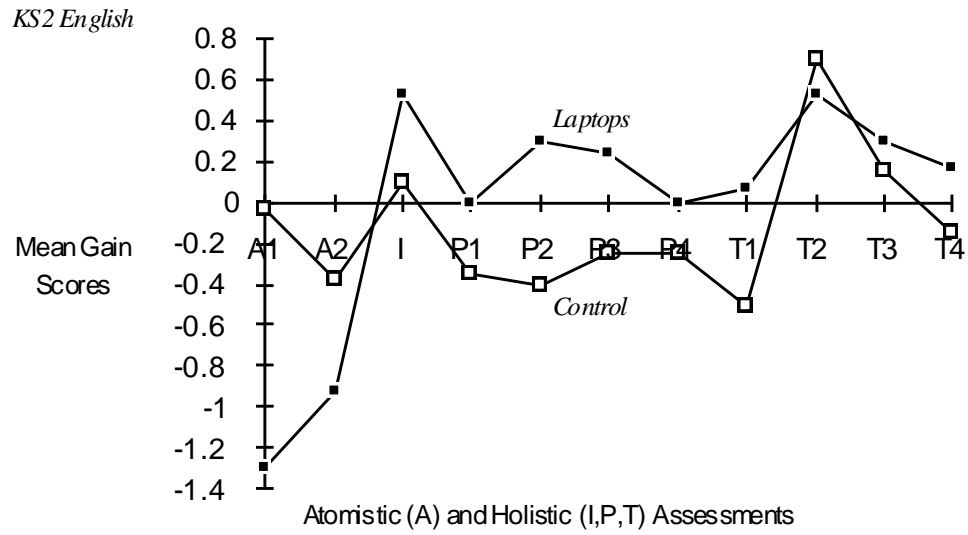**Figure 2: Gain Scores for the Matched Pairs in Science**

Figure 3:   **Gain Scores in the Atomistic and Holistic Measures for the Secondary (Key Stage 3) English Groups**

Figure 4:   Gain Scores in the Atomistic and Holistic Measures for the Key Stage 2 English Groups