Curriculum, Evaluation and Management Centre

PUBLICATION NO. 52

SUCCESS AND FAILURE IN PEER TUTORING EXPERIMENTS

C.T. FITZ-GIBBON

Explorations in Peer Tutoring
1990

2 Success and failure in peer tutoring experiments

Carol Taylor Fitz-Gibbon

Five controlled field-experiments were conducted. Two of these involved same-age rather than cross-age tutoring and neither was considered successful. The other three experiments were successful in a variety of ways; tutors generally learned more than equivalent pupils spending the same time on the content, and tutees showed significant learning gains. Two of the experiments are considered in detail, one being a replication in the UK of an experiment originally conducted in the US.

These two experiments, conducted in inner-city schools, involved 14 year olds tutoring 9 year olds in fractions. Cognitive benefits to tutors were significant in five out of six classes involved in the experiments but various confounding factors led to a cautious interpretation: at least tutors could spend considerable amounts of time helping others without falling behind in their own work.

In cross-age learning-by-tutoring projects, certain non-cognitive effects are fairly reliably present. These are the 'responsibility' effect; the 'insight-into-learning' effect; the 'teacher-empathy' effect; the 'relief-of-boredom' effect and the 'peer-tutoring-appreciation' effect.

There are many accounts of successful experiments in peer tutoring, but accounts of failures are understandably more rare. Nevertheless, if we are to understand *how* peer tutoring creates better learning and other positive benefits, and if research is to guide practice, then it is just as important to study projects which failed as it is to examine projects which have succeeded.

Of course, the dichotomy 'failure or success' is too crude. The outcomes of tutoring are multi-faceted, and there may be aspects of 'success' and aspects of 'failure' within a single project. In the main, we shall be concerned with two broad categories of outcomes: the learning of both tutors and tutees (cognitive outcomes) and their attitudes (affective outcomes).

The tutoring projects considered here are listed in Table 2.1. Some selected aspects of the Los Angeles Fractions project are summarised because, although a full account is available elsewhere (Fitz-Gibbon, 1975), this account is not readily accessible. Moreover, the method and results are needed here since the first of the projects undertaken in England ('UK Project I') was designed to replicate the major features of the Los Angeles project.

UK Project I is reported in some detail, using both qualitative and quantitative data. The purpose of this detailed examination is to cast light on *how* peer tutoring works. UK Project II has been reported elsewhere (Fitz-Gibbon and Reay, 1982) and is included here only as another example of a 'successful' cross-age project from the same research programme. UK Projects III and IV represent rather clear failures. Both were 'same-age' rather than 'cross-age' projects. An account of these projects is available in Fitz-Gibbon, 1981.

All the projects were 'learning-by-tutoring' projects in which both tutors and tutees were expected to learn the topics under study. None was a 'tutorial service' project in which tutors were providing a service to 'tutees'. In these projects, therefore, tutoring was used as a teaching strategy and pupils' participation was no more voluntary than in any other classroom activity.

The Los Angeles Fractions Project

The experiment was conducted in an inner-city junior high-school in California. The school's population was approximately 90 per cent black and 10 per cent Hispanic-American.

Tutors were 40 ninth-grade students (14 year olds) randomly selected from four low-achieving mathematics classes. Tutees were 68 nine year olds randomly selected from the fourth-grade classrooms at an adjoining elementary school. Tutoring was conducted for a three-week period and aimed at 11 objectives concerned with the addition of fractions. Immediate post-tests and questionnaires were administered, followed three months later by retention tests. The cognitive tests contained five items measuring each of the 11 objectives.

The initial training of tutors was brief and concentrated primarily on the content to be tutored: fractions. It consisted of three class periods during which the addition of fractions was explained, methods of teaching the meaning of fractions were demonstrated, and some use was made of role-play. In three of the four ninth-grade classes, tutors and non-tutors received this initial instruction together. In these three classes, the non-tutors then worked with their regular teacher on the 11 objectives throughout the next three weeks; tutors spent the same three weeks

Table 2.1 Summary of the tutoring projects

Title	Type	Topic	Tutors	Tutees	Age difference	Main design (a)
US Project (Los Angeles)	Cross-age	Fractions	9th Grade (14 yrs +) Lower half in ability	4th Grade (9 yrs +) Mixed ability	5 yrs	RO X O RO O
UK Project I (replication of US project)	Cross-age	Fractions	Drawn from two 4th form CSE Classes (14+ years old)	Two mixed-ability classes of 9 and 10 years olds	4 or 5 yrs	RO X O RO O
UK Project II	Cross-age	French: weather, numbers etc	All pupils in the lower of two 4th form French classes	All pupils in a 1st form introductory class	3 yrs	O X O O O
UK Project III	Same-age	Maths: area, bar graphs, networks, probability	Four 4th form lowest stream maths classes		0 yrs	RO X_1 O RO X_2 O RO O
UK Project IV	Same-age	Maths: solving equations	Two 4th form lower stream maths classes		0 yrs	R X_1 O R X_2 O R X_n O

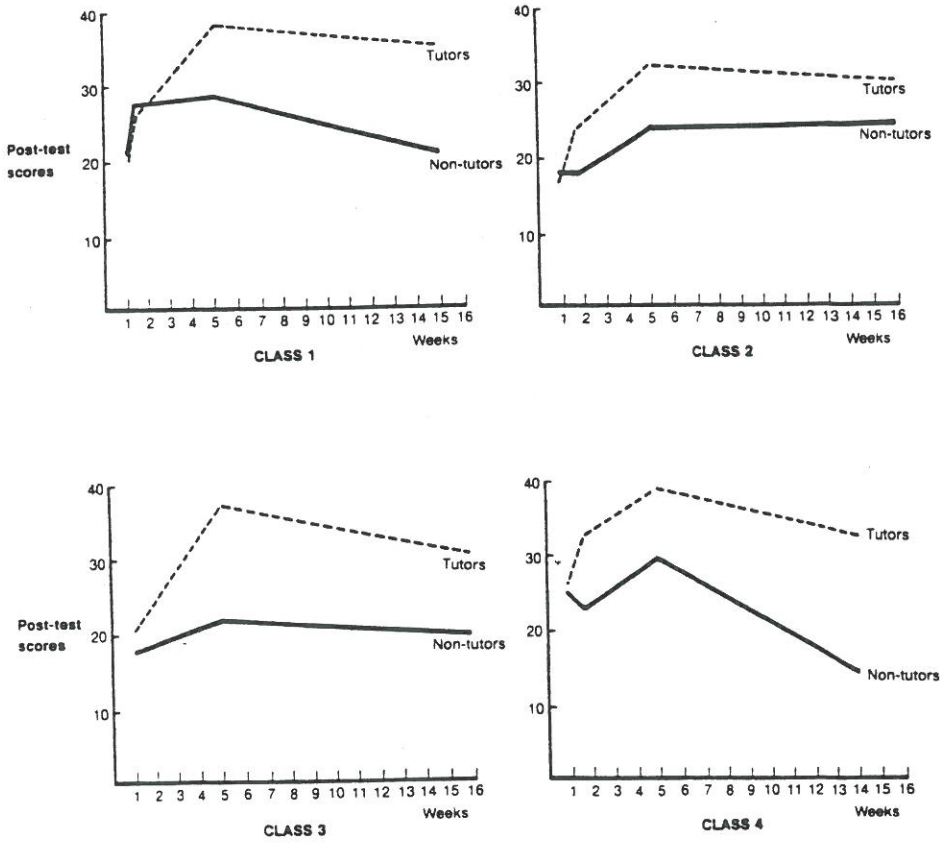
(a) Notation follows Campbell and Stanley (1966)

teaching the objectives to the nine year olds under the supervision of the experimenter. Non-tutors in those three classes thus constituted a competing-treatment control group: they received classroom instruction whereas the experimental group participated in supervised tutoring. The time allocated to supervised tutoring was either 30, 40, 60, 65 or 80 minutes per day and tutors worked with either one, two or four tutees.

In another class, the teacher had already 'done' fractions, so tutors were simply taken out of the class for initial instruction and for the three weeks of tutoring, during which time the class continued with its ninth-grade general maths programme. Since no effort was made to have the class work on the 11 objectives, students in this classroom formed a no-treatment or baseline control group.

At pre-test, in all four classes, there were no significant differences (even at the .25 level), but on the immediate post-test there were differences between tutors and non-tutors in favour of tutors (Figure 2.1). These differences remained on the retention test given three months later. The major effect seems to have been the better learning experienced by tutors in

Figure 2.1 Results of the LA Project by classroom



the course of the tutoring project (between pre- and post-test) rather than any differential retention; the eroding effects of forgetting seemed to apply to all groups.

The net result, however, was that three months after the intervention, at 'retention-test' those who had tutored knew significantly more about fractions than they had before the project; those who had not tutored had often sunk back to their pre-test levels of performance, having a set of scores on the retention-test not significantly different from their scores on the pretest. The almost total lack of retained progress seen in the non-tutor group must have occurred routinely every year, since every year fractions were dealt with in maths classes and every year, with lower-achieving pupils, teachers found it necessary to re-teach from scratch. Tutoring appeared to have broken this depressing pattern.

The mean score obtained by students in the no-treatment control group was approximately 22; the mean for the competing-treatment control group was 27. The groups consisting of tutors obtained means ranging from 33 to 43. Using Tukey's HSD, a contrast of the competing-treatment control group with the tutoring groups was significant (t = 3.02, df = 65, p = .004). Tutors who had spent three weeks teaching fourth graders about fractions had themselves learned more about fractions than had equivalent students practising fractions in their maths classes.

Means and 95 per cent confidence limits for ninth-graders on the immediate post-test are displayed in Figure 2.2.

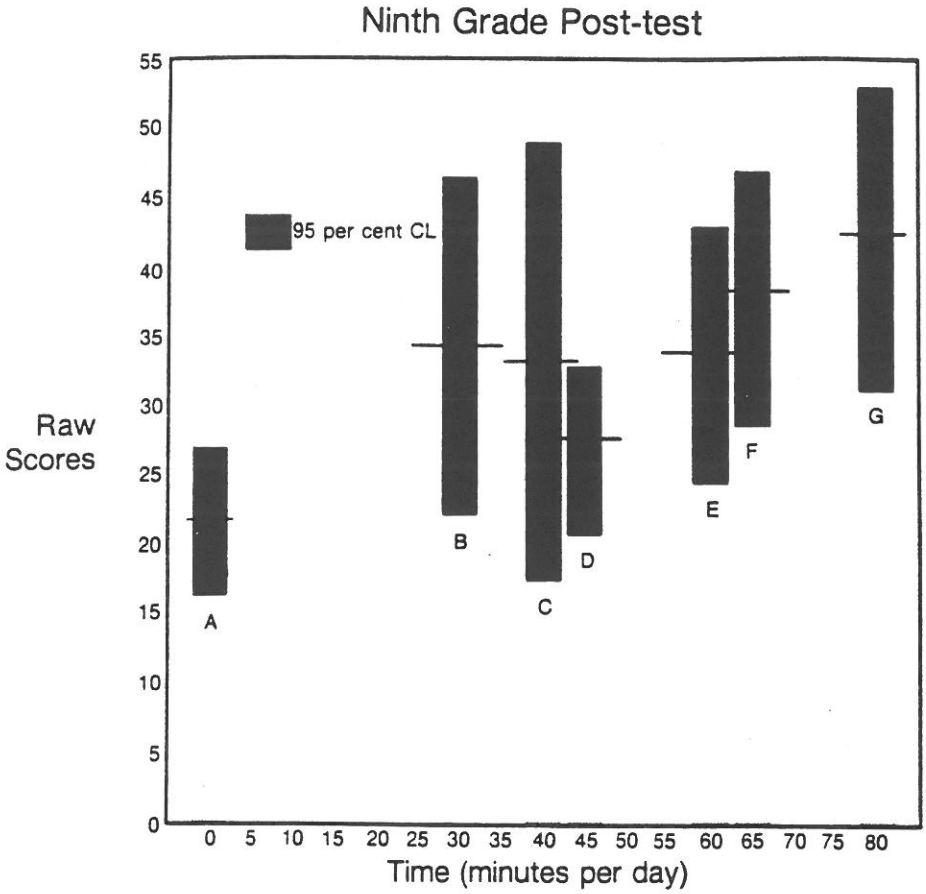
The effectiveness of low-achieving students as tutors

Had these low-achieving students been effective in teaching the fourth-graders? All fourth-graders took a post-test covering the 11 objectives but employing smaller numbers than involved on the ninth-grade test.

On this fourth-grade test, the items were randomly ordered rather than occurring in groups of five as had been the case in the ninth-grade test. This randomly-ordered test was called the 'Scrambled Fractions' test. Means and 95 percent confidence limits on the means are displayed in Figure 2.3 for the non-tutees ('0 minutes time allocated') and for tutees who had received 20, 30 or 40 minutes per day. A one-way, fixed-effects, analysis of variance on the post-test scores indicated significant effects ($F_{(3,90)}$ = p < .001). Tutees had clearly learned from the tutoring provided by the ninth-graders.

Many practitioners with experience of running tutoring programmes in schools maintain that less-able students make better tutors, perhaps because they are patient and tolerant of repetition. To examine the relative effectiveness of the tutors in this study, tutee residual-gains were computed by regression of post-test scores on three variables: amount-of-tutoring

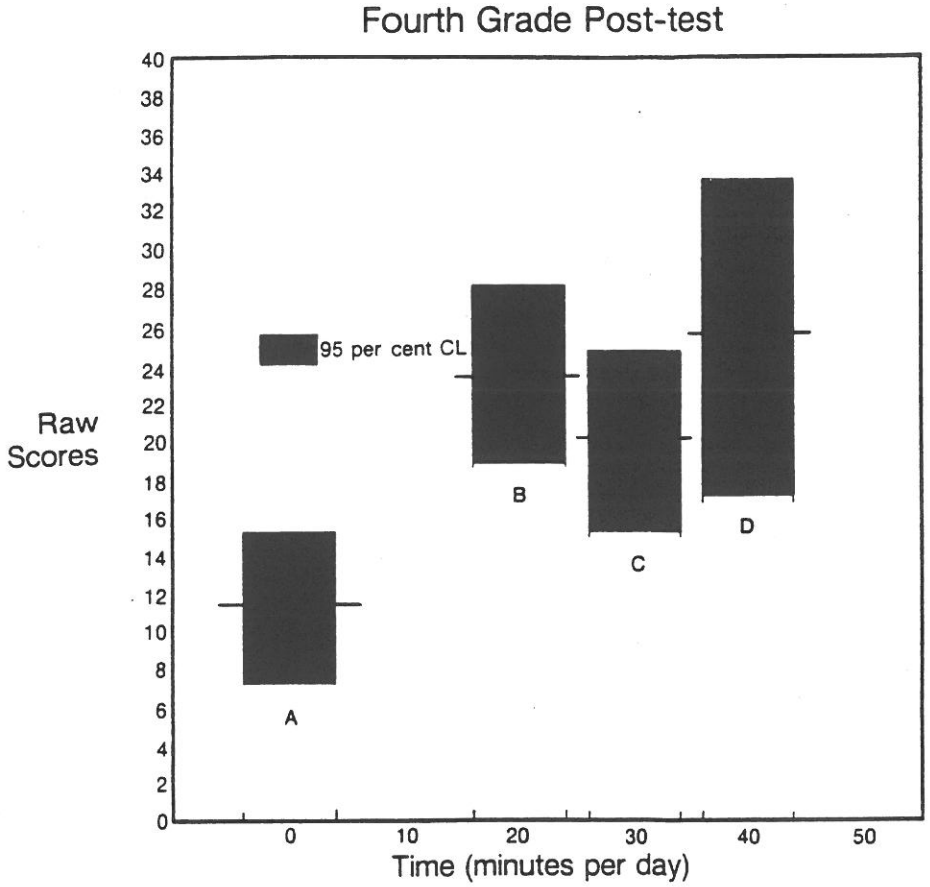
Figure 2.2 Ninth grade post-test results:
Post-test means and 95% confidence limits for ninth graders in the seven conditions.



Notes: A = no treatment control group
B = competing treatment control group
C, D, E, F, G = tutoring groups

actually received (in minutes); tutee-ability; and tutee's 'effort' characteristics. ('Effort' was a rating by the student's teacher of the student's usual level of academic effort in the regular classroom.) For each tutor, the average residual gain of his or her tutee was then computed. What kinds of tutors produced the highest achievement in tutees when time, ability, and effort were controlled? As a first indication it can be noted that tutees' average residual gains were not significantly correlated with tutor-ability ($r = -.09$), pre-test ($r = -.14$) or post-test ($r = -.15$). The correlation between tutee residual gains and tutors' retention test scores was -.28, significant at the .05 level and there was a significant negative correlation between average tutee residual-gains and tutors'-effort ratings ($r = -.30$,

Figure 2.3 Fourth grade post test-results:
Post-test means and 95% confidence limits for fourth graders allocated zero, 20, 30 or 40 minutes of tutoring daily.



Notes: A = no treatment control group
B, C, D = tutored groups (tutees)

p = .03, n = 39). These trends indicated that low-effort, low-achieving students were slightly more effective as tutors than were high-effort, high-achieving students. The tendency was not a strong one but *was* sufficient to refute any suggestion of a tendency in the opposite direction for this group of students (a group, it must be remembered, of generally low achievement). These findings are consistent with the observations of many practitioners: poor students are frequently relatively effective tutors.

How can persons less competent in the subject be as effective as, or more effective than, more competent persons when acting as tutors? In considering answers, it must first be recalled that *tutoring is not teaching.* Tutoring means explaining a limited, usually prescribed, set of concepts; very often it is performed following training. Had tutors had responsibility for

selecting the curriculum, analysing it, and organising it into clear objectives, then one might expect subject-matter competence to become important, if not crucial, to success. But in the context of clearly-delimited instruction, and with training, low-achievers can make effective tutors.

Furthermore, one possible explanation as to the slightly greater effectiveness of less-able pupils is that these pupils covered the objectives slowly and laboriously – which helped the tutees to learn. In other words, lower-ability tutors avoid the mistake of presenting material too quickly. This explanation found support in another finding of this study: tutors who underestimated the amount their tutees had learned had tutees with higher residual-gains (a reversal of the commonly-cited, though not empirically well-supported, expectancy effect, cf Rosenthal and Jacobsen, 1968; Elashoff and Snow, 1971).

Another indication of how less-able tutors may be particularly effective may be deduced from reports such as those by Stallings and Kaskowitz (1974) and Soar (1972). They presented evidence from large-scale analyses that teacher's use of open-ended questions, ie questions high in the Bloom taxonomy (Bloom, 1956), was negatively related to achievement. Direct questions at lower levels of the Bloom taxonomy facilitated acquisition of knowledge by students. Lower-achieving or less-competent tutors may be likely to ask simple, direct questions rather than indirect, abstract, reasoning-type questions.

Retention tests

Tests three months after the brief tutoring experiment showed tutors still with a significant advantage over non-tutors. Non-tutors, in fact, had sunk back to a mean score not significantly higher than their pre-test score. Among fourth-graders, tutees still showed significantly higher average scores than non-tutees ($t = 2.68$, $p<.05$. Effect size = .81).

Summary

In summary, as far as cognitive outcomes were concerned, the results of this experiment conducted in a school setting indicated that after 14-year-old pupils had received initial instruction, acting as a tutor produced significantly higher achievement than did classroom-instruction. Nine-year-old tutees made significant gains as a result of receiving tutoring and both tutors and non-tutors still showed significant advantages over equivalent control groups three months after the three-week experiment. There was a slight tendency for low-effort, low-achieving students to have been the more effective tutors.

The LA project had the following characteristics:

- Cross-age with an age-gap of five years (14-year-old pupils tutoring 9-year-old pupils).

- Cross-school: tutors left their secondary school to tutor in a primary school.
- A structured set of objectives which tutors were assigned to teach, in the area of fractions.
- Tutors were low achieving: tutees a complete range of the school's intake.
- Initial training consisted of three days in the regular classroom, tutors along with non-tutors, explaining the objectives and including some role-playing of tutoring: these lessons conducted by the researcher.
- The tutoring sessions were supervised by the researcher; the normal classroom-lessons were conducted by the regular teachers who, because of the removal of half of the class for tutoring, benefited from a reduced class-size.

The last-named characteristic represented a flaw in the design, the kind of flaw which one must often accept due to practical problems involved in setting up field experiments. The effects of tutoring could have been attributed to the effect of the supervising teacher. Efforts were made in the replication study to aviod this confounding of teacher with treatment.

UK Project I Cross-age (Fractions)

This was the project which most closely resembled the inner-city LA project.

Participants

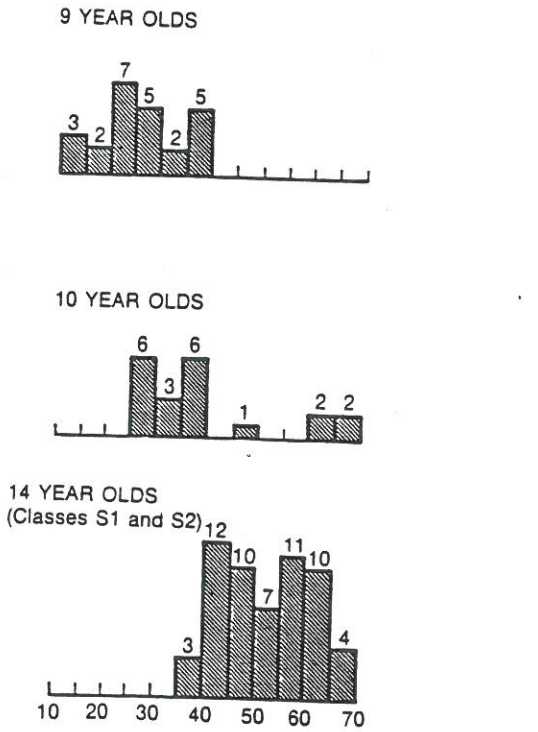
A secondary school was located which had a primary school nearby. It was a comprehensive school serving pupils in a lower socioeconomic status area in north-east England. With the agreement of the teachers of the classes involved, the Head of Mathematics chose two classes in the fourth form to participate in the project. These classes were randomly divided into tutors and non-tutors; they will be referred to as class S1 and class S2 (secondary 1 and secondary 2). The two classes were at the second level in a four-level stream and were preparing for CSE examinations. For comparison purposes, two classes from the level below the experimental classes (S3 and S4) were also given the pre-and post-tests.

At the primary school, two classes of nine- and ten-year-old children represented the entire range of ability for the school's intake. Since the number of tutees needed was almost equal to the number of pupils in these classes, there was no control group for the tutees. The class containing children of ten years and older will be referred to as P1; the class with children of 9 years and older as class P2.

Design

The project was designed to examine two major effects: the effect on secondary pupils of being assigned a tutoring role as opposed to remaining in a regular classroom situation; and the effect on tutors of tutoring one tutee for 30 minutes as opposed to two tutees consecutively for 15 minutes each. In other words, there were two manipulated variables in the design: tutoring vs not-tutoring and, within the tutoring condition, having one as opposed to two tutees. The latter planned variation would also allow tutees to receive different amounts of the treatment in so far as some tutees received 15 minutes of instruction on each occasion whereas others received 30 minutes. The design was implemented as a pre-test post-test 'true' experiment (Campbell and Stanley, 1966) with secondary pupils assigned randomly to the tutoring- or no-tutoring condition, and the tutors then randomly assigned one or two tutees. The assignment of the tutees to the tutors was based on the rank-ordering of both groups on the pre-test, with higher-scoring tutors assigned higher-scoring tutees. Such a procedure was considered necessary to avoid the unsatisfactory situation of a tutee knowing more than a tutor. Despite the age difference of about 5 years, it was apparent from pre-test distributions that this situation could easily have arisen: 16 of the 44 primary children tested fell within the range of the tutors' scores (see Figure 2.4). It must be noted that the tutees covered the

Figure 2.4 Distribution of scores on the fractions pre-test



entire ability range for the school's intake, whereas tutors did not include any pupils from the top stream.

Procedures

The timetables at the two schools dictated the extent to which tutoring could be implemented. For one experimental class (S1), tutoring could occur during two out of the three weekly one-hour lessons, making a total of six tutoring sessions in the three weeks of the project. For the other experimental class at the secondary school (S2), scheduled activities at the primary school (such as radio broadcasts, visits to the swimming-baths) meant that tutoring could only occur on one of the three lessons per week, yielding therefore only three tutoring sessions during the three weeks of the project.

A major problem in the design of the LA experiment had been the confounding of supervision with experimental condition – the researcher supervised the tutors throughout the experiment, and some of the strong positive effects could have been attributed to the effect of the supervising teacher rather than just to the effect of tutoring. It was possible for class S1 in this experiment to control more effectively for the supervising teacher effect: the Researcher Associate and the S1 teacher alternated between taking the normal class instruction and supervising the tutoring. Unfortunately this balance was not achieved in S2, in which the regular teacher supervised only one of the three tutoring sessions and the Research Associate worked with the tutors the rest of the time.

During lessons when tutoring was scheduled, the non-tutors worked in a normal classroom fashion consisting of demonstration of the problems on the board, discussion and exercises. For tutors, the procedure was first to assemble and walk across to the primary school (this took about ten minutes). At the primary school a room was ready with tables-for-two placed irregularly throughout the space available. Tutees were brought in for tutoring for 30 minutes, the remainder of the one hour period being taken up with completing 'Tutor Record Sheets' and returning to the secondary school. On the 'Tutor Record Sheets', tutors were to tick off which objectives they had worked on and write comments on how the session had gone.

Instruments

The objectives to be covered during the planned three weeks of the project were agreed upon by the teachers and the Research Associate. A fractions test was constructed, with five items per objective and this was administered in the primary and secondary classrooms two or three weeks before the project started. The Standard Progressive Matrices (SPM) test (Raven, 1958) was also administered to all classes by the Research Associate. Due to the danger of a ceiling effect on the fractions test, (some had scored as high

as 70/72 ie 97%), a more difficult test was constructed for the post-test. A questionnaire was administered to tutors at the end of the project.

Table 2.2 shows reliabilities for the pre- and post-tests used in this project with internal consistency measures ranging from .71 to .91.

Table 2.3 summarises the experimental conditions and the pre-test scores. As would be expected from the random assignment procedure, tutor and non-tutor groups were equivalent on the Standard Progressive Matrices and on the fractions test. Differences were not even statistically significant at the .20 level.

Table 2.2 Reliabilities of the Objectives Based Test

Sample	n	Internal Consistency[a]		Test-retest	
		Pre-test	Post-test	r	p
Tutees	39	.87	.91	.85	<.001
Tutors	25	.73	.71	.37[b]	.04
Non-tutor	63	.74	.86	.65	<.001
All tested	127	na	na	.81	<.001

Notes: (a) Cronbach's alpha

(b) The low test-retest correlation for tutors suggested that tutors' rank orders changed considerably from the pretest. For class S1 the test-retest correlation was .72 whereas for S2 it was .42. This suggests the re-ordering of ranks was primarily in class S2. This re-ordering appparently brought achievement more in line with ability since the correlation with the SPM increased from .41 to .66. In class S1 correlation with ability changed only marginally, from the surprisingly low values of .16 to .10.

Implementation and pupil response

In the following paragraphs, comments from pupils and teachers are used to illustrate both how the project was implemented and how pupils responded to it.

Attendance

Attendance was good, as is often the case in the first term of a school year. There were no substantial or significant differences between the attendance of tutors (average of 93%) and non-tutors (average of 90%) as measured by their presence in maths classes.

Teachers' reports

Teachers were asked to record their impressions of the project in writing. The Head of the primary school, where the tutoring actually took place, wrote:

Table 2.3 Experimental conditions and pre-testing results

Class[(a)]	Stream	Experimental condition	Periods allocated: To Tutoring	Tutor Preparation	Normal Classwork	n	SPM: $\bar{X}$	SD	Fractions Pre-test: $\bar{X}$	SD
S1	2	Tutors	6	3	0	13	45	6.3	53	8.3
		Tutor control	0	0	9	14	45	8.3	54	10.1
S2	2	Tutors	3	6	0	13	45	4.9	52	8.9
		Tutor control	0	0	9	15	43	7.0	54	8.1
S3	3	Comparison Gp.	0	0	9	18	43	6.8	45	8.9
S4	3	Comparison Gp.	0	0	9	17	41	8.7	48	5.8
			Tutoring time allocated							
P1	Mixed ability	Tutees (10 yrs +)	Either 3 × 15 mins = 45 or 3 × 30 mins = 90			20	38	10.5	42	14.0
P2	Mixed ability	Tutees (9 yrs +)	or 6 × 15 mins = 90 or 6 × 30 mins = 180			24	35	8.0	26	8.2

(a) S = Secondary P = Primary

> When I had the opportunity to visit the library where the children were working, I was most impressed by the quiet, controlled working atmosphere. The junior children treated the older girls and boys with respect (and most of them referred to the secondary school children as 'students') and the secondary children were patient and pleasant with them.

One of the two secondary teachers involved observed:

> Attitude and interest of tutors seem to be generally positive (varying in enthusiasm from pupil to pupil of course). Tutees, in general, seem to be keen and more than happy at being taught in this way.
>
> (Teacher of class S1)

The other teacher wrote:

> The tutors responded in a very adult and responsible way. They worked hard in class and on their files, were concerned with their tutees' progress as well their own, and have shown considerable subsequent interest in the exercise.
>
> (Teacher of class S2)

Pupil reports

The written comments from pupils provided a glimpse of their attitudes to, and perceptions of, the project. Some provided accounts of what happened between tutor and tutees.

Written comments are a very difficult kind of data to handle. Inevitably, a certain amount of 'exampling' is used: one makes a point and selects a quotation which illustrates it. This is, of course, highly-selective use of the raw data.

In trying to move towards some summary of this kind of data, one strategy is to form categories and count the occurrence of comments which fall into these categories, preferably checking reliability by having the assignment to categories cross-validated by other raters. But breaking up of a set of comments from one person often loses what might be termed the 'clinical-insight' which is gained by considering all the comments from one 'case' together, as a miniature case-study.

The problem is illustrated by the case shown in Figure 2.5. Tension and ambivalence are discernible, pervading a stressful experience of tutoring for a not-very-able pupil. This pupil, incidentally, made an above average residual gain during the project.

Using categories, this pupil would have contributed counts to such categories as: 'being influenced towards teaching' (lines 19, 20) 'being influenced away from teaching' (lines 5, 6), 'thinking the project was good' (lines 13, 14, 16, 17, 18) 'liking it' (line 5) and 'hating it' (lines 4 and 11). The approach of counting categories was not used.

Figure 2.5 One tutor's comments

Time comment was written	Comment	Line No
Prior to tutoring:	I'm looking forward to it bit nervus	1
After tutoring on Day 1	He no's about fraction's. did not have very long. Bite nervise A bite slow	2 3 4
Day 2	I like it. Got to know him a bit more. I would not wan't to become a teacher.	5 6
Day 3	I need to work on my fractions. Any tutor has to work on his fractions.	7 8
Day 4	I like it but getting a bit sick of it. I know him much more.	9 10
Day 6	I enjoyed today because it was the last one. I am sick of keep coming over here. It help't me to do fractions much better than before. He nows more than he started with I think. It would be good if we could do this again in the 5 year to help us to revise for are exams. I would like to teach children between 9–11.	11 12 13 14 15 16 17 18 19 20

Tutees' comments

In reporting the comments from tutees, an indication will be given of the characteristics of the pupil who made the comment. 'Relative gain', one of the characteristics to be reported, comes from 'residual gain analysis'. In a residual gain analysis, the pupil's post-test performance is interpreted in the light of the score which would have been expected from that pupil, knowing his or her pre-test score, Standard Progressive Matrics' (SPM) score, and the general gain made by the class as a whole. Thus the residual gain score shows the learning gain made by the pupil *relative* to other similar pupils in his or her class. These relative gains were classified from high to low in five groups, the terms used for reporting the gains being those shown in Table 2.4.

In the presentation of comments below, tutees are described by five pieces of information:

Class	Age-group of class	Sex	Entry level	Relative gain

Table 2.4 Categories used for reporting relative gains made during the project

Standardised residual gain score (Relative gain)	Descriptive term applied
Over 1.5	very good
0.5 to 1.5	good
-0.5 to 0.5	average
-0.5 to -1.5	below average
below -1.5	poor

The 'entry level' was a composite standardised score for each class, based on the fractions pre-test and the SPM score. As with relative gain, pupils' relative standing in their class has been expressed on a five-point scale, this time using the descriptive categories A, B, C, D and E. An 'A' indicates a pupil at the top of his or her class.

Take, for example, this pupil:

P2: 9 years, Female, C, average gain

This pupil was in class P2 (Primary 2), of nine years olds. She was about average for her class ('C') and made gains during the project which were about average for pupils with similar scores on the fractions pre-test and on the SPM.

Here, from the pens of tutees, are some descriptions of tutor behaviours:

> I thought my teacher was very nice. She understood when I needed help. She was also very patient, she also explained everything very clearly. The things I did I thought were *just* a bit hard, and when I forgot to cancel the fractions down she just told me I'd forgotten and said remember next time.
>
> (P1: 10 years, Female, C, Good gain)

> My teacher was nice. He told me to ask him if I was stuck and he said 'Here's a ruler and rubber'. He explained all right but he said 'do it my way' and I did not understand his way so I got them wrong. I think the work was hard because I could not understand the sums.
>
> (P2: 9 years, Female, C, Good gain)

> I think my teacher was very nice. I think the scheme was a good idea, because when we get to the secondary school it will save the teachers a lot of time teaching us how to do it (the fractions). She explained things very well and she was very understanding. Some of the fractions I already knew but some of them I didn't know. Most of them were very simple. I enjoyed it a lot.
>
> (P1: 10 years, Female, C, Good gain)

> I like my teacher very much, she helped me and I think I knew a lot more about fractions when I finished than when I started. She was very good at teaching but she was a little bit impatient and she told me to hurry up. When I had just started, she said "Are you finished yet" but she was funny. She did some sums for me. And she said she didn't like doing fractions very much but she wanted to be a teacher. I thought the work was just right.
>
> (P2: 9 years, Female, C, Average gain)

Some of the younger tutees in particular found the work hard and thought the tutors spoke too quickly.

> I thought my teacher was kind. But she said the words too fast.
>
> (P2: 9 years, Male, B, Average gain)

> I thought my teacher was talking a little too fast. She didn't explain properly and she kept making mistakes herself. She was very nice though. She helped me as much as she could and she did some sums for me. I thought the work was a little hard for me. I didn't learn anything because it was too hard. I thought the idea was a very good idea even though I didn't learn anything. I think it might have been easier for the teacher because she didn't have as many children in the class when we went for fractions.
>
> (P2: 9 years, Female, C, Poor gain)

> I thought that my teacher was a bit too fast. When I was doing a sum which was hard he said "hurry up, you're taking a long time." He taught me for a half hour. When I was there for a while I started to do fractions OK.
>
> (P2: 9 years, Female, C, Average gain)

> My teacher was a bit too quick because she told me the answer when I could have got it if I was given a minute.
>
> (P2: Male, B, Average gain)

As might be expected, since tutees were drawn from a full ability-range across two age-groups, there were also tutees who found the work easy. Apparently they enjoyed the experience nevertheless:

> The teacher was nice. It was a good idea. I liked it. It was organised well. I had a patient teacher. I didn't learn anything but it was good practice.
>
> (P1: 10 years, Male, A, Good gain)

> It was good and the kids that taught us didn't rush us. The sums were dead easy and simple. We did those sums last year.
>
> (P1: 10 years, Male, B, Average gain)

> My teacher was very nice. When I got there each day we used to have

> sweets. We had to hide them from [the Research Associate] because you are not allowed to eat sweets. If he had found out that we had sweets we would have to chuck them in the bin. I liked my work I was doing. I thought it was easy work and he was always telling me the answers and I got all of them right.
>
> (P2: 9 years, Male, D, Scored zero on post-test)

> [Perhaps always telling a tutee the answer was not good tutoring. But for this below-average nine year old the post-test was too difficult, so interpretation is problematic.]

Tutees had opinions about the factors influencing tutors' effectiveness as teachers.

> My teacher was a good one because he did one thing at a time.
>
> (P2: 9 years, Male, B, Average gain)

Many tutees expressed the thought that it was a good idea to have pupils as tutors:

> I think it was good because they were younger than other teachers and I think they understood you better.
>
> (P2: 9 years, Male, B, Average gain)

Tutors' comments

Turning from tutee reports to tutor reports, a further impression of the process of tutoring was gained from examination of tutor record sheets. Below, comments from the tutors have been grouped into postulated 'effects'. In the writer's experience, most cross-age tutoring projects yield anecdotal evidence of these effects and while it is unlikely that fundamental or substantial readjustments in attitudes could be expected from a few brief experiences of being a tutor, one might wonder what the cumulative effect would be of regular use of periods of cross-age tutoring.

Tutors are described by five pieces of information:

Class	Sex	Entry level	Relative gain	Tutee's relative gain

It will be recalled that some tutors had two tutees.

The responsibility effect

Teachers often comment with pleasure and surprise at the effort tutors make, the responsibility they seem to feel towards their tutees. One aspect of this effect was the attitude revealed when, before the first tutoring-session, those who were to be tutors were asked to write down what they thought of the idea. Their comments frequently reflected anxiety as to whether they would be able to tutor well.

> I feel a bit worried in case I get the fractions wrong, but other wise I am quite looking forward to it.
>
> (S1 Female, C, Below average gain. Tutee gain: good)

> I'm looking forward to it but nervous.
>
> (S1 Male, D, Good gain. Tutee gain: good)

> It might be very embarrassing for you if you forget how to do the sums yourself.
>
> (S1 Male, B, Below average gain. Tutee gain: below average)

In low achieving secondary school mathematics classes, it is sometimes quite difficult to induce in pupils any sense of *needing* to learn mathematics. Prim references to later employability or examination success mean little to restless teenagers. The comments just cited indicated the way in which the tutoring project provided tutors with an *immediate* need to know the work.

Tutors' daily records often reflected their concern and the efforts they were making to get tutees to understand.

> . . . at times it could be frustrating when they did not understand the work.
>
> (S1 Female, B, Below-average gain. Tutee-gain: below average)

And the elation clearly felt by one tutor after the tutoring session:

> Success! Claire has worked hard this morning and got the hang of No. 9 straight away.
>
> (S1 Female, D, Average-gain. Tutee-gain: average)

> I sometimes feel terrible when he is just sitting there and I think he's stuck so I go to help him and he says "I know" and carries on.
> (S1 Female, C, Good-gain. Gain made by tutee referred to: average. Other tutee: good)

The insight-into-learning effect

Early proselytisers of peer tutoring claimed that tutoring would cause pupils to 'learn how to learn'. (Gartner, Kohler and Riessman, 1971). Whilst no empirical evidence for this effect appears to have been presented, it remains an interesting hypothesis. The hypothesis receives some support from comments like that below:

> I think that this project has been quite useful to me. It has helped me notice how much you must concentrate and listen to the teacher's explanations of how to do the sums and how important it is to ask if you get stuck.
>
> (S1 Female, B, Below average gain. Tutee gain: below average)

Teacher empathy effect

Being cast in the role of tutor could be expected to induce some understanding of teachers' roles.

> I think I have learnt how to teach and I found out how hard it is for a teacher to teach 30 kids never mind 1 or 2.
>
> (S1 Male, B, Below average gain. Tutee gain: average)

> I didn't really learn anything apart from teachers must have a hard job.
>
> (S1 Male, C, Below average gain. Tutee gain: below average)

The relief of boredom effect

Tutoring was a valuable experience if only in that it was a change from routine; the boredom that is too often felt at school was held at bay. It must be stressed that this relief was not gained at the expense of basic cognitive instruction. Tutors *were* working on fractions but some nevertheless regarded the project as a relief from normal schooling.

> It makes a change from being in the same classroom and being taught by the same old teacher.
>
> (S1 Male, B, Below average gain. Tutee-gain: below average)

A tutor, one of whose tutees showed one of the lowest residual gains, wrote in clear, neat handwriting prior to tutoring:

> I am looking forward to teaching these children as it means getting out of the classroom as I do not like maths but I will like teaching these children.
>
> (S1 Male, C, Average gain. Tutees' gains: poor and below average)

Another, more effective, tutor wrote

> It has been great not just because missing maths in school, it is very exciting.
>
> (S2 Male, C, Average-gain. Tutees' gains: very good and average)

The peer tutoring appreciation effect

Both tutors and tutees perceived some benefits from their being more close in age to each other than are regular teachers and pupils:

> Also tutees you teach might feel better with their own generation teaching them.
>
> (S2 Male, C, Good gain. Tutees' gains: average and poor)

Some tutors appeared to be aware of the idea that a reason for teaching work might be to assist their own learning. Thus one comment was:

> It would be worth continuing because it saves some teachers a lot of time explaining when they already know.
>
> (S2 Male, C, Average gain. Tutees' gains: very good and average)

and an unintentionally humorous comment:

It helps the tutee to learn as well as teaching your self. This could be improved but it is a very useful way of teaching young kids of 9 and 10. Instead of having fully growning adults doing it . . .

(S2 Male, C, Good gain. Tutees' gains: average and poor)

Cognitive outcomes

Did tutoring have a positive effect on the achievement of tutors? A no-treatment control group was provided by the pupils in classes S3 and S4, the undisturbed other two fourth-year groups, with a correction applied for different initial levels of performance. The results of this comparison, not surprisingly, showed statistically significant effects ($p < .001$) in favour of tutors. Tutors scored on average 46 on the fractions post-test, as against an average of 29. Adjusted for differences in initial level, the mean would be 44 as opposed to 31, still a difference of 0.98 of a standard deviation. That is to say, the effect size for a comparison with a no-treatment group was 0.98. The comparison of tutors with no-treatment control classes only answered the question; 'Did tutors gain from the tutoring project?' They did. But the more important question is whether tutors gained more from tutoring than they would have gained by spending the same time on the same topics in the regular classroom. For this question, the appropriate comparison was with the pupils in classes S1 and S2 who did not tutor but who worked on the same topics. This 'competing-treatment' control group will be referred to as the 'non-tutors'.

A straightforward initial test, pooling results from classes S1 and S2 is shown in Table 2.5.

Table 2.5 Simple comparison of tutors vs non-tutors on the immediate post-test

Group	n	$\bar{x}$	SD	t	p (one tail)	Effect size
Tutors	25	46 (67%)	8.6	1.08	.14	0.25
Non-tutors	30	43 (62%)	13.9			

Although the tutor average was slightly higher than that of the non-tutors, it did not look as though the strong effects found in the Los Angeles experiment had been replicated. A difference significant at the .04 level could be shown, however, if simple gain scores (post-test minus pre-test) were used. This was due to the fact that the random selection had by chance favoured the non-tutors (cf table 2.3). Both these results were sufficiently positive to maintain that the tutors, who had spent from three

to six hours (out of nine hours of maths lessons) on helping primary school pupils had done at least as well on the post-test as an equivalent group of pupils remaining in a regular classroom situation. *Tutors had not provided tutoring at the expense of their own learning.* Furthermore, it must be remembered that the class size in the regular classroom situation was reduced by about 50% due to the withdrawal of the tutors. For non-tutors the pupil-teacher ratio was 1:13 and 1:18 in S1 and S2 respectively, whereas it was 1:26 in the tutoring sessions when the tutees were being supervised as well as the tutors.

To some extent, effects weaker than those found in the LA experiment had been expected for a number of reasons. There was a tighter design: the confounding of tutoring with supervision by an outside person had been removed in one experimental class (S1). There was also a considerably smaller amount of 'treatment': the amount of actual tutoring had been only either 180 minutes (class S1) or 90 minutes (class S2) in the three week period, whereas the amount had ranged from 280 to 1120 minutes in the Los Angeles experiment. Third, there was a different context: it was felt that the regular classroom (the control group condition) was more effective in this school than in the stressed conditions of a Los Angeles 'ghetto' school. Thus the difference between the effectiveness of tutoring in comparison with the control condition would be expected to be less, since the 'control' was more effective.

Given that tutors did not apparently show any decrement in achievement, and did enjoy and possibly benefit in other ways from the experience, it might be argued that this was sufficient evidence to justify continuing to recommend cross-age tutoring as an instructional procedure. But such a position might fail to take sufficient account of the fact that organising cross-age tutoring requires a considerable amount of extra work from teachers. One would much prefer to see cognitive as well as non-cognitive benefits. Further analyses of the data were undertaken in an attempt to elucidate the conditions under which tutoring seems to be beneficial for the tutors.

The graphs in Figure 2.6, overleaf, show the post-test scores separately for classes S1 and S2 and for boys and girls in those classes. Only for females in class S1 was the mean score for non-tutors higher than that for tutors. Separate analyses (Table 2.6) showed that the differences due to tutoring/non-tutoring in class S1 were not statistically significant, whereas those in class S2 were, at the .02 level, one-tail. Thus having been picked by the random-assignment process to tutor resulted on average in no significant difference in class S1 but significantly better performance in class S2.

Clearly it is important to consider the characteristics of the two classes and differences in the way the project was implemented. These contextual factors are summarised in Table 2.7 on page 49, along with some cognitive and non-cognitive outcomes.

Figure 2.6 Post-test scores for S1 and S2

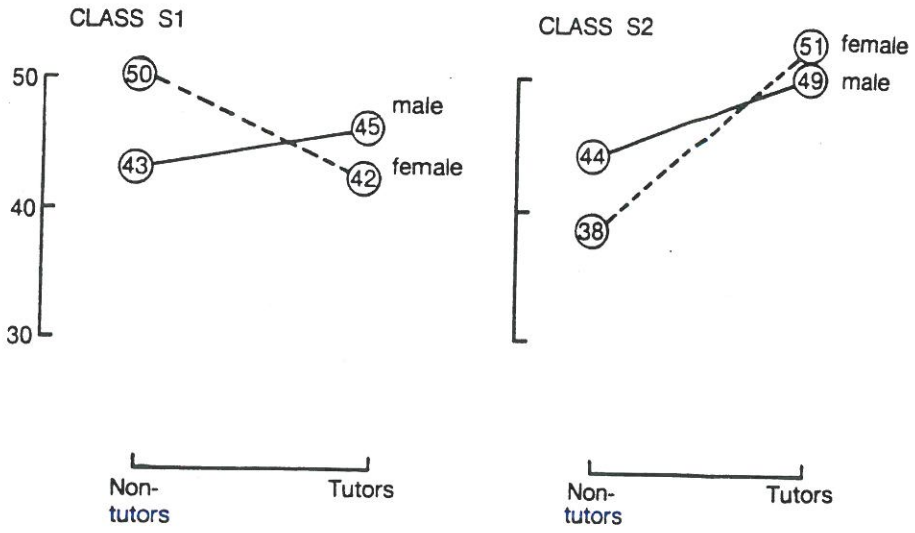


Table 2.6 ANCOVAs for post-test scores

Class	Source	df	MS	F	P
S1	Sex (S)	1	7.68	0.121	0.73
	Tutoring (T)	1	35.24	0.553	0.47
	S x T	1	4.42	0.069	0.79
	Covariates:				
	SPM	1	0.91	0.014	0.91
	Fractions	1	1220.09	19.154	0.01
	Residual	17	63.70		
	Total	22	108.79		
S2	Sex (S)	1	129.74	1.608	0.06
	Tutoring (T)	1	376.79	4.670	0.04
	S x T	1	39.06	0.484	0.49
	Covariates:				
	SPM	1	1097.85	13.607	0.01
	Fractions	1	85.66	1.062	0.31
	Residual	21	80.68		
	Total	26			

Table 2.7 Summary of Project I cross-age (Fractions)

	Class S1	Class S2
Characteristics of class		
Teacher attitude to tutoring	doubtful	neutral to positive
Pupil co-operation level in normal classes	high	not as high
SPM achievement correlation	.16 (n = 23)	.46 (n = 27)
sex achievement correlation	.001	.13
Implementation		
Supervision of tutors by research associate	5/9 periods	8/9 periods
Time allocated to objectives:		
Tutors	9 lessons	9 lessons
Non-tutors	about 7 lessons	9 lessons
Number of tutoring sessions	6	33
Number of sessions:		
before tutoring	1	1
preparation during tutoring	3	6
Class size for supervision:		
tutors and tutees (a)	average of 22	average of 17
non-tutors	13	18
Cognitive outcomes		
Post-test: tutors	62%	72%
non-tutors	65%	59%
Retention test: tutors	68%	57%
non-tutors	68%	54%
Tutor attitudes		
Per cent positive response to 'Would you like to have tutoring as a term's course?'	62%	100%[b] or 85%
'Would you like your whole maths class to go regularly to tutor?'	38%	82%

(a) Class size for supervision was 26 when tutors were tutoring tutees, 13 where preparing for tutoring. The averages arise from different proportions of tutoring and preparation.
(b) Only 11 of the 13 tutors answered the questionnaire.

Complexities in a seemingly simple experiment

Teacher attitude

There is a school of thought which claims that the outcome of any innovation depends crucially on the attitudes of the teacher who implements it. It would be fair to report that the teacher of class S1 was less enthusiastic about tutoring than the Research Associate or indeed than the teacher of class S2. He said, for example, 'I wouldn't want it to be thought they learned more by going and tutoring than by being in my class'. His somewhat negative view of tutoring was one possible factor in the differences in outcomes. However, there were many other factors and this one should not be given undue weight. There is no systematic evidence regarding teacher effects on the outcomes of experimental implementations (trials) of innovations, effects which would surely depend in any case on the extent to which teacher actions affected the particular innovation. In the present case, it was felt that the experiment was given a fair chance in both classes concerned and that differences in the attitudes of the teachers should not be seen as a major explanation of the cognitive outcomes, although teachers' attitudes may have influenced tutors' attitudes.

Pupil co-operation level

If, as is often claimed, tutoring wins co-operation from pupils who are otherwise poorly motivated and indifferent or even antagonistic to work in regular classrooms, then it follows logically that the beneficial effects of tutoring when compared with regular classwork would be greater the more difficult the pupils. This hypothesis could be invoked to explain the different effects in classes S1 and S2 and also the lesser effects in this project when compared with the LA project in which general classroom conditions were much worse.

Correlations

The correlations of SPM and of sex with achievement are listed for interest and future reference rather than because their interpretation is yet clear. The lack of correlation between sex and achievement suggested that in each class boys and girls were achieving equally. This was not the case in all classes. The class of ten years olds in the primary school had a correlation of $-.45$ (females were achieving somewhat less well in fractions than males) and the secondary school class S3 showed a correlation of .33, a weak trend in the opposite direction. The correlation of .46 ($P < .05$) in class S2 between achievement and the SPM, a measure of aptitude, is difficult to interpret; but this correlation may be an indicator of classroom processes, one of the possible statistical indicators about which more research is needed (Lohnes, 1972). However, not for nothing have

correlation coefficients been called the slippery statistics, the *difference* between the correlation in class S1 and that in class S2 was not statistically significant ($z = 1.11$, $p = .28$).

Supervision

Supervision was balanced in class S1 between the Research Associate and the teacher, whereas in class S2 the Research Associate supervised the tutors for eight of the nine mathematics lessons. Attempting to control for the effects of a supervising teacher is one of the major problems facing experimentation in schools, since teacher effects might not themselves be stable and might vary with the situation. Certainly, however, the alternation of supervision achieved in S1 was a tighter design than the confounding of supervision with treatment that occurred in Class S2. Consequently it must be noted that some of the positive effects of tutoring found in class S2 might have been due to the supervision of tutors by the Research Associate.

However, this finding is not without quite positive implications for tutoring as a technique. As in the LA experiment, we have the situation of a stranger coming into a fairly difficult situation and getting a high level of co-operation and hard work from the pupils. Would the same level of co-operation have been afforded a stranger who came in to continue regular classwork? Supply teaching in the inner city is not known to be an easy assignment for a teacher, even in the UK. In Los Angeles, it was perceived as inviting disaster unless you were an intimidating male of the husky footballer variety: yet a small female with a foreign accent (ie English) was able to run a three-week tutoring project with no trouble. In short *tutoring may be valuable in situations where pupil co-operation is a problem.*

Time allocated to objectives

Under this heading, note has been taken of the fact that the non-tutors in class S1 finished the objectives before the end of the project and moved on to other work. This could represent a cost of tutoring projects: they may slow down coverage of the curriculum topics. Again, this reinforces the concern to make sure that tutoring projects are run as effectively as possible, as far as cognitive learning is concerned. If topics are covered more slowly, they must be learned more thoroughly – if, that is, we judge the value of a project by what is learned academically. It could be argued that teaching fewer topics in a way which might promote co-operation and caring would be better than getting through the syllabus regardless of pupils' attitudes to others.

Number of tutoring sessions and training sessions

As already noted, class S1 had six periods in which they could tutor whereas class S2 had only three. Because class S2 could not tutor so often, they had more time to prepare for tutoring during the project (two lessons of

preparation to one lesson of tutoring per week). This 'mix' might be another factor which led to the greater success of tutoring in Class 2. Again, we confront one of the complexities of classroom experimentation: a treatment must be evaluated not as an isolated event but also as an event in a sequence of events. The effectiveness of the procedure might depend upon events immediately preceding or following the procedure, the 'mix' of experiences. The importance of a mix of procedures can be inferred from another controlled experiment in classroom settings. Chalip and Chalip (1978), ran a five-week experiment in which three sets of pupils were assigned to work in three different ways:

(1) Co-operatively in small groups
(2) Individually
(3) In an alternating fashion, cooperatively one day and individually the next.

No affective measures or sociometric choice measures showed significant differences, but on a test based on the learning objectives (noun and verb identification), the mixed condition produced significantly higher achievement ($F(2.29) = 12.50$, $p < .001$). Interestingly, however, better achievement for the mixed-treatment groups did not occur in a similar experiment in which the mix of co-operative and individual working conditions was achieved by a switch half-way through the experimental period, rather than by daily alternation (French et al 1977). These results seem to tell us that not only is variety the spice of life, but it should be peppered in, not introduced in lumps. As for tutoring, careful consideration must be given to the balance between tutoring and preparation for it during a project. It must be mentioned in this connection that the Research Associate reported that the tutors themselves were disgruntled at having to spend time preparing for tutoring when they would have preferred to be tutoring. Teachers may have to schedule the preparation lessons needed, ignoring opposition from tutors (to be realistic, however, the balance may well be determined, as it was here, by the timetable rather than by educational judgements).

Class size

The effect of class size, or the pupil-teacher ratio, during the project has already been mentioned; the point made was that it favoured the non-tutors, thus providing an even more stringent test of the effectiveness of tutoring.

Summary

The two classes differed in the effects of the tutoring project. In one class, achievement as measured by the post-test showed no significant differences, while in the other, tutors achieved more highly. These differences in outcomes could be attributed to differences in implementation procedures

and/or to contextual effects such as teacher attitudes. Only further experimentation will elucidate the effects of tutoring. At least tutors did not suffer any immediate or long-term decline in achievement as a result of spending some time helping primary pupils.

One or two tutees?

It will be recalled that some tutors, randomly selected, were assigned one tutee to whom to provide 30 minutes of tutoring; others were assigned two tutees whom they tutored for 15 minutes each, consecutively.

What might one expect? That longer tuition would lead to higher tutee-gains perhaps? Or would the same amount of instruction merely expand to fill the time available? Tutors might cover the same ground more slowly, leading to better retention or they might socialise a little – leading, perhaps, to the more positive effect associated with longer sessions. Or might tutors run out of work and feel embarrassed? Would tutors themselves benefit from repeating the lesson to a second tutee?

It was fairly clear from reading their comments about tutoring that many tutors who had been assigned two tutees had not found the short 15-minute sessions satisfactory:

> The first tutee got on very well with his starting out his shapes but there is one thing: that is you don't get long enough with your tutees. When you have two kids to teach it is difficult to get all of the things you would like to do in the small amount of time.

This less satisfactory experience had perhaps slightly affected their attitudes, as shown on a questionnaire. There was a tendency for tutors who had been assigned two tutees to show less enthusiasm towards the idea of regularly tutoring in mathematics classes than tutors who had had one tutee.

Even tutors with only one tutee found the time short:

> "I think we should be given more than half an hour"
>
> (S2 -12: Female, B, Average gain. Tutee missed post-test)

> "The project could be improved by having more time with the child. . . ."
>
> (S2 Female, C, good gain. Tutee gain average)

One tutor who felt 15 minute sessions were too short for the tutees seemed to have formed the hypothesis that since having two tutees meant teaching each objective twice, this might enhance the tutor's learning:

> "It would be much better if we just had one pupil each because with two pupils you can not get on, but it helps yourself to learn more about fractions".
>
> (S1 Male, D, Average-gain. Tutee-gain: one absent, one poor)

However, comparison of post-test and retention test scores did not yield much support for the hypothesis of better tutor learning with two tutees. Post-test means were 46.7 for the 11 tutors with one tutee and 45.5 for the 14 tutors with two tutees, an entirely non-significant difference ($p = .84$). However, one might expect that the repetition of the tutoring might have led to better long-term retention of the topic. Retention tests given 14 weeks after the end of the project did show that tutors in each class who had had two tutees scored higher than tutors who had had only one tutee. Again, however, the effects were weak and not statistically significant (in class S1, 59% as opposed to 56%; in S2, 55% as opposed to 45%- $F(1,24) = .52$ $p = .46$)

Cognitive effects, then, were lacking, but written comments from several tutors indicated that they preferred longer tutoring sessions and therefore one, rather than two, tutees. On the Tutor Questionnaire, those with two tutees, while not reporting that they themselves found the sessions too short, tended to think their tutee found the session 'short' whereas the average response from tutors with one tutee indicated that the tutee found the sessions 'OK', a difference which was statistically significant ($t = 2.04$ $p = .05$). However, overall, the shorter sessions seemed to have had no effect on tutors' reports of their own or their tutees' enjoyment of the project nor on the amount of preparation for tutoring sessions they reported having done at home.

Long-term effects

Cognitive

The fractions post-test was re-administered as a retention-test some three months later. As reported in Table 2.7, scores at this time showed that in each class tutors and non-tutors were indistinguishable; but class S2 showed more decline than class S1, perhaps due to differences in work covered between the two tests. So we are confronted with evidence that, unlike the LA project in which cognitive benefits from having been a tutor were still in evidence three months after the project, in this project the most positive comment that can be made is that tutors did not seem to have suffered from having spent time helping younger pupils. Taking time out to tutor younger pupils might have been a positive experience as far as enjoyment and attitudes were concerned and it had not detracted from the tutors' own learning.

However, the retention test was a measure of achievement *in the topics which were involved in the tutoring project.* How about achievement in other parts of the mathematics syllabus? Were there any unanticipated effects there? The following report from the teacher of S2 was disturbing:

> Many of the tutors were placed in the bottom half of the class on the after-Christmas exam – quite a proportion of which was based on the exercise material. I'm not sure of the significance of this at this stage.

Following this report, the school exam papers were borrowed for analysis. This revealed that the difference between tutors and non-tutors would not normally count as 'statistically significant', reaching only the .20 level. Nevertheless, there was a difference of 10 points between the means on a test of about 60 items and this was a cause for concern. It seemed possible that there was an effect which might be termed a 'post-Hawthorne depression'.

Following the interesting and novel experience of tutoring, the regular classwork might have seemed dull in comparison and have been less well attended to. If this had been the case, the school tests referred to above in the teacher's report should have shown tutors doing less well on the work which was covered in class but the same or better on work covered in the tutoring project. Analysis of the test items revealed that this was indeed the case. The entire difference was accounted for by the difference on items which had been covered in class, not in the tutoring project.

The observations here are not conclusive enough to suggest that there was definitely a post-Hawthorne depression leading to lower achievement on other parts of the curriculum; but it does raise an interesting issue in evaluation : How would we value a project that was so pleasant that it made regular classwork particularly unappealing by contrast? One response would be to recommend that the project be continued, replacing regular classwork.

However, perhaps the solution lies in the notion of variety referred to earlier. Continued longer, tutoring too could satiate and regular classwork become the welcome change. Regular breaks from tutoring to renew enthusiasm for it have been strongly recommended (Fitz-Gibbon, 1978: 39) and a meta-analysis of 65 projects indicated maximum cognitive benefits had been derived from projects lasting four weeks or less (Cohen, Kulik and Kulik, 1982).

Non-cognitive effects

The same teacher who reported the problem with the school-test scores continued:

> Nevertheless, I hasten to recommend the exercise because I feel the benefits far outweight the disadvantages.

And in the same report he recorded his perception of the benefits:

> Most of the children have matured because of the experience and have a more realistic attitude to learning and a more sympathetic attitude to teachers. They have a confidence in themselves – having been teachers – but tend to recognise their own difficulties in learning. Of course there are exceptions to this.

In responding to the Teacher Questionnaire item: 'Were there any pupils for whom you felt tutoring was particularly useful? If so what kind of pupil and why?' The teacher wrote

> For a term or so the lads who used to cause trouble or be lazy and truant.

These perceptions of some of the longer-term effects of a tutoring project, added to the evidence for non-cognitive benefits presented earlier, leave the impression that positive effects on various tutors' attitudes are fairly reliably produced and are perceived both by tutors and by teachers. These attitudinal effects are often highly valued and perhaps account for the great enthusiasm for tutoring shown by many practitioners. Experimentation, however, suggests that claims regarding *cognitive* benefits need careful scrutiny. Perhaps the greater ease of measuring cognitive gains makes our standards of evaluation in this area very much more strict; but it might equally be the case that cognitive benefits are not as stable an effect of tutoring as are the immediately visible effects of the change of role. Indeed, in general, given equal opportunities to learn, classroom processes may have more effect on pupils' attitudes than on pupils' cognitive achievements. This may well be true of tutoring: it may have more effect on pupils' attitudes than on their learning.

Conclusion

In this chapter, controlled field experiments have been reported from four inner-city Los Angeles classes and two inner-city classes in north-east England. All experiments involved low-achieving 14 year olds tutoring nine or ten year olds in fractions.

The method of the controlled experiment was chosen in order to obtain stronger inferences than would be possible with less rigorous methods. However, while the interpretations of controlled experiments are far less fraught with uncertainties and ambiguities than are interpretations of 'passive observational' studies (Cook and Campbell, 1979), nevertheless results never determine the conclusions which can be drawn. They only *put constraints* on those conclusions.

These constraints are often sobering for any claimed 'better method'. Experiments keep researchers from believing their own sincere rhetoric. In the present instance, as mentioned briefly at the beginning of this paper, two *same-age* projects were not successful. In the *cross-age* projects, the significant gains demonstrable in the four LA classes were only replicated in one of the two UK classes, and even there the advantage to tutors was no longer detectable on a retention test 14 weeks later, whereas in LA the benefits had apparently been retained. Many possible explanations can be

considered for the difference between the US and UK results. In particular, one suspects that normal classroom instruction – the control group's treatment – was more effective in the UK. Also, perhaps, one must consider the much shorter experimental period in the UK and the fewer resources committed. The LA project ran for three weeks and employed a room for tutoring with special booths and practical equipment such as Cuisennaire rods, circles of paper, scissors etc, whereas the UK project was run for only two weeks in a spare classroom in which tutoring pairs sometimes tended to distract each other.

On the other hand, considering other kinds of learning than academic learning, confidence has been increased that the *cross-age* tutoring role will evoke in tutors strong feelings of responsibility towards tutees, insights into the learning process, expressions of empathy with teachers, relief-from-boredom, a recognition that peer tutors may be able to assist learning and, very importantly, high levels of co-operation with their own teacher or with a stranger.

These are valuable effects and their value will become even more appreciated if schools come to face the kinds of social adjustment problems experienced in inner-city Los Angeles schools. New ways become more valuable when traditional ways no longer work.

Acknowledgements
The support of the Social Science Research Council (now the Economic and Social Research Council) is gratefully acknowledged. The author also wishes to record appreciation for the energy and skills which David Reay brought to the UK projects.

Correspondence
Dr C Fitz-Gibbon, School of Education, University of Newcastle-upon-Tyne, St Thomas Street, Newcastle-upon-Tyne, NEI 7RU.