

EVALUATION REPORT

EV511

THE ROLE AND DESIGN OF BASELINE STUDIES IN THE EVALUATION OF ENGLISH LANGUAGE TEACHER TRAINING IN THE CASE OF GUINEA.

BY

E HAWTHORNE - C LANE

February 1994

CONTENTS

[Preface](#) - [Abbreviations](#) - [The Study](#) - [Overall Conclusions](#) - [Main Findings](#) -
[Lessons to be Learned](#) - [The Study](#) - [The Pre- and In- Service Teacher Education Project](#) -
[Introduction](#) - [Inputs and Costs](#) - [Personnel](#) - [The Sample](#) - [Methodology](#) - [Problems During](#)
[Implementation](#) - [Results](#) - [Main Findings of the Study.](#)

ANNEXES

[Study Documentation](#) - [Chronology of Study](#) - [Table A1: Numbers of Students](#) - [Table A2:
Gap-Filling](#)
[Test Results](#) - [Table A3: Dictation Test Results](#) - [Table A4: Writing Test Results](#) - [Table A5:
Oral Test](#)

Results - Table A6a: Percentages of Respondants Concerning the Importance of Training Programme Elements - Table A6b: Percentages of Respondants Concerning the Sufficiency of Training Programme Elements - Table A7a: Number of Responses Concerning the Balance within the Training Programme and the ENS Methods of Training - Table A7b: Number of Responses Concerning the Usefulness of Resources and the Bo Seminar.

PREFACE

Each year the Overseas Development Administration (ODA) commissions a number of ex post evaluation studies. The purpose of the ODA's evaluation programme is to examine rigorously the implementation and impact of selected past projects and to generate the lessons learned from them so that these can be applied to current and future projects.

The ODA's Evaluation Department is independent of ODA's spending divisions and reports direct to the ODA's Principal Finance Officer.

Evaluation teams consist of an appropriate blend of specialist skills and are normally made up of a mixture of in-house staff, who are fully conversant with ODA's procedures, and independent external consultants, who bring a fresh perspective to the subject matter.

This study was conducted by Dr Cyril Weir, a language testing specialist of the University of Reading, and Mr John Burton of Evaluation Department assisted in the completion of this final report.

This study involved the following stages:-

- initial desk study of all relevant papers;
- consultation with individuals and organisations concerned with the project, including field missions to collect data and interview those involved;
- preparation of a draft report which was circulated for comment to the individuals and organisations most closely concerned;
- submission of the draft report to the ODA Principal Finance Officer, to agree the main conclusions and lessons to be learned from the study on the basis of the draft report;
- agreement with the evaluator on the final report, which is published together with a summary sheet (EVSUM).

This process is designed to ensure the production of a high quality report which draws out all the lessons.

J C H Morris

Head, Evaluation Department

ABBREVIATIONS

CALS Centre for Applied Language Studies

ELT	English Language Teaching
EFSA	English for French-Speaking Africa
ENS	Ecole Normale Superieure de Maneah
ISSEM	Institut Superieur des Sciences de l'Education de Maneah
KELT	Key English Language Teacher
ODA	Overseas Development Administration
PGCE	Postgraduate Certificate of Education
PTT	Pupil Talking-Time
TC/TCO	Technical Cooperation/Technical Cooperation Officer
TEFL	Teaching of English as a Foreign Language
TTT	Teacher Talking-Time
TT	Trained Teacher
UT	Untrained Teacher

OVERVIEW

1. THE STUDY

1. The ODA's Evaluation Department has undertaken two studies on the effectiveness of English Language Teacher Training, the first in Nepal and the second in Guinea. As initially conceived, the Nepal study set out to generate a method for the interim and ex post evaluation of ELTT projects by establishing the nature of the baseline data required for such evaluation; to develop methodologies for the collection of such data; and to collect these data for the projects in question. That study was thus a vehicle for the development of methods and methodologies. The subsequent study in Guinea serviced to test and build on the findings of the Nepal study. Each study, therefore, generated lessons both for the design and implementation of baseline studies in future, and for the design and implementation of ELTT projects.

2. The study was undertaken during phase II of the overall English Language Teaching (ELT) Project in Pre-service and In-service Teacher Education that had been launched in 1985. It set out to establish a baseline against which to evaluate the effectiveness of a pre-service English as a Foreign Language teacher-training course in Guinea-Conakry in francophone West Africa as the first cohort of teachers from that programme went into the schools. The study was undertaken by UK consultant Dr C.J. Weir with the assistance of a local research team in Guinea. It was funded by the Evaluation Department of ODA. The study took place from 1990-1992 and this is its final report.

3. The methodology of the study was to identify and collect baseline data that would illuminate the impact of the teacher-training project on the performance of students. The study, which was a nonequivalent control group/experimental group, pre-test/post-test design, sought to establish whether the training had any impact on pupils' language scores over a two year period. The control group was opportunistic and consisted of the 14 remaining teachers in Guinea who had not received

any training. The experimental group was composed of those trained teachers who managed to secure a position in the secondary system. A test battery, including a written part and an oral part, was administered. The first administration was carried out at the beginning of the first year of the study; the second at the end of the same year, or, in some cases, at the beginning of the next; and the third at the end of the second year.

4. The test battery, which was based on the textbook "English for French-Speaking Africa" (EFSA 6^{eme}), had four components: reading (gap-filling), writing, listening (dictation), and an oral. It was designed to assess teacher performance by using pupils' output as an indicator of learner intake.

5. Observational data were obtained in order to provide evidence on whether, after training, the experimental group of teachers faithfully implemented their training when appointed to language teaching posts in Guinea. The observational instrument was designed to collect data on the following: (a) teacher verbal interaction with pupils in English and French during fixed coding periods; (b) the teaching and learning activities taking place during the whole lesson; (c) presence or absence of certain criterial features in the conduct of lessons; (d) the relative amount of time devoted to teacher and pupil talk, in English and French, in the whole lesson.

6. Three series of visits to schools were conducted during two academic years (1990-1992). The data reported in this research cover only the observations carried out on those teachers who were present at all of the three series of visits. The data consist of 51 observations (33 Experimental and 18 Control). Each teacher who remained in the study to the end (11 experimental and 6 control) was observed three times during that period.

7. A tracer study of former trainees was employed with the intention of providing data for use in the setting up of a new one-year postgraduate course in education at the English section of the Institut Supérieur des Sciences de l'Education de Maneah (ISSEM) which would replace the original four-year programme. It consisted of a questionnaire in two sections: the first sought graduate teachers' ratings of the original course components, and the second their comments, opinions, and suggestions on various aspects of the course. The experimental group, consisting of fourteen graduate teachers from the programme who had succeeded in being placed in the secondary schools, received copies of the final questionnaire at the beginning of their second year of teaching experience. The completed questionnaires were handed in or sent to the research team a year later for analysis.

OVERALL CONCLUSION

8. The overall conclusion of this study was that there was no significant difference between the impact on the performance of the students of the trained, but new and inexperienced, teachers and that of the students of the untrained, but experienced, teachers. The comparatively short timescale of the study, however, and the limited amount of teaching possible in a time of civil unrest, might have contributed to the lack of any statistically significant difference in impact on the performance of the two student groups.

9. It was not possible to select inexperienced teachers for the control group because of the limited number of teachers remaining in the country who had not received any training five years into the project. Had this been possible significant differences between the two groups may well have emerged.

10. At the time, the project in Guinea was the only one available to ODA for further investigation of

the methodological procedures first applied in Nepal. Despite the known problems of establishing a suitable control group, the willingness of the project team and the recipient country to test these procedures further were seen to offer an important opportunity for such investigation.

11. Although the data suggest no statistically significant differences between the scores of the two groups of pupils at the start of the study, and in terms of language improvement over the two years of study, the raw scores of the experimental group do suggest a slight trend in favour of this group by the end. It appeared that even after a short period in the classroom, the inexperienced but trained teachers were achieving results which, at the very least, were comparable with those of experienced but untrained teachers.

12. The analysis of the observation of the teachers reveals significant differences between the trained and untrained groups in terms of the teaching practice adopted.

13. The data generated by the tracer study need to be treated with some caution, given the small sample involved. Despite a good degree of perceived satisfaction with the Maneah four-year course, in terms of the value of the components included, the teaching methods of the staff and the use of resources, the tracer study revealed a number of problems which need to be considered if the future one-year PGCE course is to be successful and effective. The study highlighted some specific areas where improvement is necessary, indicating, for example, that the amount of methodological training participants had received was insufficient.

14. The study also demonstrated the crucial inter-relationship between research and training, since feedback from ISSEM-trained teachers is to continue to be monitored. Furthermore, building on the results of the tracer study, the project staff are seeking to respond to the need to synchronise the teacher-training programme at tertiary level with the ELT curriculum at Lycee level.

15. The baseline study described in this report has been shown to be feasible even in a country with such very limited communications networks as Guinea. From the results obtained, the procedures employed in the study would for the most part appear to be effective and transferable to similar projects. Such studies would be cheaper if conducted by project staff with some external supervision. They could contribute to effective projectisation and should be initiated as close to a project's start date as is feasible. Where difficulties arise in forming a suitable control group, alternative baseline designs are available; one example involves measuring student performance in the two experimental groups against a defined standard of skills required for the labour market or the educational system.

16. Experience in Guinea demonstrates that it is far cheaper to use local project staff to conduct such studies, although this may lead to some reduction in professional rigour if monitoring by professionally qualified staff is insufficient. The involvement of counterparts in such studies provides them with a valuable learning experience in research methodology. Having counterparts write up the results of a study for possible publication provides them with a strongly motivating and enriching experience which may make an important contribution both to project sustainability and to the enhancement of their professional self-respect and commitment.

MAIN FINDINGS

Observational Data

17. From a quantitative point of view, there is no statistically significant difference between Experimental and Control groups in the overall mean percentages of language use by pupils in the

classroom in English and in French or by teachers in English and in French.

Checklist Data: Criterial Pedagogical Features

18. During the three visits, observations were made of teachers' classroom practice, using a checklist of criterial features which the trainers considered to be indicators of training success. The "checklist" data, summarising the observed indicators, show a statistically significant difference between the teachers in the Experimental group and those in the Control group on all the indicators except one (Pupils did guided writing not in the book).

19. Observational data thus suggested that trained teachers scored an average of 78% in terms of using teaching practices defined as characteristic of a trained teacher, compared to an average score of 50% for the untrained but experienced teacher (statistically significant at the 1% level).

20. The occurrence of five indicators among the Experimental group of teachers is more than 30% higher than for the Control group (see Para 2.7.8). It is particularly in these basic pedagogical techniques that training is seen to make a difference, despite the gap in the number of years of experience between Experimental and Control group teachers. In respect of these criterial features the training programme can be seen to have benefitted the Experimental group.

21. On indicator 2 there is a relatively high, and very similar, frequency of occurrence in both groups (Teacher gave model in English i.e. word/sentence). High frequencies of occurrence occur also in respect of indicators 1, 3 and 11 in the control group (though lower than in the experimental one) and attest to the fact that good experienced teachers will naturally, in the course of time, reflect certain of these criterial features in their pedagogy. Indicators 4 and 12, however, show that experience in itself will not endow teachers with all desirable techniques and practices.

22. The indicators occurring least frequently in both groups appear to be 5 (Comprehension questions asked in English i.e. - new ones not in book), 6 (Pupil produced original sentence in English - at least 3 different pupils), 7 (Teacher gave extended listening practice), 8 (Pupils did guided writing not in book) and 10 (Extra practice - e.g. own drawings/questions). These areas perhaps demand most of teachers' communication skills and are obviously in need of greater attention in the new teacher-training programme if they are to feature more in the Guinean classroom. Given that they were considered at the start of the study, to be important distinguishing features of the trained teacher their absence from teaching practice is serious. Further investigation needs to be done into the reasons behind this. The designers of the new postgraduate training programme need to establish whether their absence is due to contextual constraints in the school situation or to the fact that these features were not sufficiently well grounded in the original Maneah 4 year training programme.

23. The data from Sections 1 and 4 of the observations also suggest that the trained group did not vary greatly from the untrained group in the amount of pupil, rather than teacher, talking-time in class. The expectation of the trainers was that in the trained teachers' classes there would a better balance and that the pupils would have the opportunity to talk more.

Test Data

24. In general the results indicate significant improvement in test scores for both experimental and control groups in those students sampled from visit 1 to visit 2. Smaller changes are evident from visit 2 to visit 3, being significant only for the gap-filling tests in both groups, and for the writing tests in the experimental group.

25. Test scores were analysed using the General Linear Models (GLIM) procedure to take into

account the other, potentially relevant, variables on which there were complete data. This made it likely that any reported differences in the groups' scores was due to the treatment effect and not to extraneous variables. This analysis revealed no significant differences between the scores of the control and experimental groups in any of the tests; nor did the other variables (numbers in class, number of books, language ability of the teachers) have any significant effect.

26. The two groups' figures for improvement in performance showed no significant difference. There was, however, a non-significant but slightly greater improvement in performance on the part of the experimental group by the end of the second year of the study, when the results are compared with those obtaining at the end of year one (visit 3-visit 2).

27. Over the two year period (visit 1-visit 3) the students sampled in the experimental group showed more improvement than those in the control group in the dictation and gap-filling tests but not in the writing tasks.

28. The tests used in the study all exhibited high reliability coefficients and respectable discrimination indices. The items would form the basis of a reliable set of achievement tests for Book 6 of EFSA.

Tracer Study Data

29. Only 8 completed questionnaires from the group of 14 experimental teachers were received and this obviously represents only a very restricted sample of the current members of the teaching force trained solely in Guinea. A wider sample is necessary before there can be any confidence in the general applicability of the data.

30. The limited data indicated a general satisfaction with the nature of the course components and the way they were taught but there was evidence that participants felt that more time and attention could have been paid to many of the pedagogical elements of the course. There is an assumption here that the trainees would be aware of deficiencies in their own training. If the pedagogical demands placed on them in the school situation are minimal then any such deficiencies might not be apparent or serious.

LESSONS TO BE LEARNED

Lesson 1. An educational impact study with a baseline component is the most appropriate means with which to show the effect of different training inputs, in terms of performance gains from a baseline point, as the baseline study is able to isolate the impact of different levels of ability prior to training.

Lesson 2. Such studies may be most useful where a project is ongoing as they can then inform subsequent project phases and the cost, relative to the project cost, would be lower. They may also be justified in pilot projects which are likely to be replicated or expanded in the future.

Lesson 3. The results of a baseline study are most meaningful if the measured gains made can be assessed against a benchmark set by local needs. In this way a benchmark language proficiency target could be set for each situation, for example for entry to further study or to meet employment demands; this would enable evaluators to form a judgement on the benefits generated.

Lesson 4. An analysis of the role of English in the economy should precede any substantial ELT project and baseline study.

Lesson 5. If local institutions are involved in the conduct of a study there is a danger of bias and personal interest, and the financial incentives may be insufficient to guarantee a study's systematic

implementation. The involvement of an outside, disinterested, professional group is a safeguard and external moderation lends credence to the study's results.

Lesson 6. Local staff can gain useful training and research experience through involvement in the conduct of such studies; this also helps to reduce costs. Adequate supervision of such staff is, however, required to avoid a reduction in rigour and the introduction of bias.

Lesson 7. Given the likelihood of attrition of participants over time, sample sizes should be large enough at the outset to sustain the study through its length. The longer the period of the study, the larger the sample groups should be. The logistical implications of this, however, must not be overlooked.

Lesson 8. Where attrition of participants in control groups is likely to be high, alternative baseline designs should be considered. In place of a control group, it may be appropriate to consider a comparison of the experimental group with an average standard of performance, determined by randomly testing all the current students, to establish benchmarks of performance on exit from each year. The design of such instruments could facilitate the monitoring of national standards in language performance in the subsequent period.

Lesson 9. The use of language test data to monitor gain in student language achievement also provides training benefits. Their use can provide clear lines of development for the testing of language at all levels within a school system.

Lesson 10. The design process underlying the construction of a range of valid, reliable and practical tests also requires project staff to be very clear about project objectives. In the long term such tests could provide useful data where the existing school leaving examinations are at odds with the objectives of the intended training.

Lesson 11. The observation checklist, the observer training manual and the use of self report data, offer a tried methodology for monitoring teacher-training projects, and could be used for clarifying project objectives in feasibility studies.

Lesson 12. It is necessary to obtain, through observation, data upon teacher practices. Teacher-training projects should specify the measurable changes in teachers' classroom performance that will be an outcome of the training. This is best done by establishing a trained teacher profile which includes a criterial list of features to differentiate trained from untrained teachers. An evaluation should demonstrate a relationship between student scores gains and the training provided. If there is no difference between the two groups (control and experimental) in terms of the test data it is important to establish why. In such a case the observational data can help establish whether there was a problem with the training and indicate the areas in which this manifested itself.

Lesson 13. The difficulties in setting up and monitoring educational baseline studies in countries with poor communication facilities should not be underestimated. Poor communication facilities in Guinea (unreliable postal and telecommunications networks, lack of a resident Embassy or British Council Office) certainly undermined the prompt collection of data and contributed to the small number of questionnaire returns. In such cases a member of the project team needs to be designated as in-country co-ordinator, to assist in overcoming these difficulties.

MAIN REPORT

1. BACKGROUND

1.1. THE STUDY

1.1.1. The shift in 1984 from a socialist to a liberal type of government in Guinea-Conakry led to changes in the educational policy of the republic of Guinea in general, and to the language policy in particular. As a result of these changes the decision was taken to reintroduce English into the secondary school curriculum. English language was made a compulsory part of secondary education. The great demand for teachers of English which these changes entailed could not be met in the normal way; the usual qualification for a teacher of English in the secondary system was a university degree, but there were simply not enough graduates to satisfy demand. As a stop-gap measure, untrained, unqualified teachers were employed, but this was clearly not a situation which could be allowed to last.

1.1.2. The training of English teachers to meet the needs of secondary schools became a priority. In cooperation with the British government, an English department was set up in October 1985 at the Ecole Normale Supérieure de Manéah (ENS) and a four-year pre-service ELT teacher-training course was developed. The ENS was an advanced teacher-training college situated in Manéah, about 43 km from Conakry, the capital city. Although the college was created in 1979, it had no English department until the need to reintroduce English as a subject in the secondary school curriculum emerged. The ENS did not run postgraduate courses as its intake was recruited from among students successfully completing their second year of undergraduate studies. In all departments, except for the English department, the training was designed to last three years in addition to the first two years at university. It aimed in the first place at imparting in-depth knowledge of the different subjects to be taught at secondary level, namely Mathematics, Physics, Social and Natural Sciences. The skill-getting training - in other words, actual learning about teaching - would start only towards the last year of training.

1.1.3. In the specific case of the newly created English department, the course lasted four years and intake consisted of secondary school leavers with no background in English but with enough basic concepts in science to register for further studies. These would-be teachers appeared to be much younger than their counterparts in other departments and stayed on a year longer than the latter. Their course had two main components: language instruction and methodology in English as a Foreign Language but the emphasis was laid on the former, with only language being taught in the first two years.

1.1.4. Thirty students out of the initial intake graduated four years later in June 1989 and, eventually, most of them were able to secure teaching posts in the secondary school system.

1.1.5. The study reported on in this paper set out to evaluate the effectiveness of the first version of this teacher-training course. In other words, did the course achieve effective training of English teachers for the secondary system? The study was a nonequivalent control group/experimental group, pretest, post-test design. Observations were to be used to provide data on whether, after training, an experimental group of language teachers faithfully implemented their training when they went into the language teaching classrooms in Guinea. Test data were used to establish whether the training had any impact on student language scores over a two year period.

1.1.6. These investigations were planned to start in October 1989 but could not take off until December 1990 because the graduates were not assigned to posts until the second half of the year following their graduation. In that same period the decision was taken, as part of the new educational policy, to convert the former ENS into the Institut Supérieur des Sciences de l'Éducation de Manéah (ISSEM). In accordance with the new decision the ISSEM, unlike the former ENS with its emphasis

on academic rather than professional training, has been assigned the mission of designing and running exclusively postgraduate and tailored practical courses in education in order to meet the growing and changing needs of secondary school students and personnel. Training courses would no longer last more than a year unless otherwise indicated.

1.1.7. The English Section (no longer a department) now needed to design a one-year postgraduate course in education with a special focus on methodology. It is to this end that a tracer study was added to the main study to establish retrospectively the opinions of graduate students concerning the value of various elements of the training course they had undertaken. Thus the originally intended primary emphasis of the study upon summative evaluation shifted somewhat to include a formative dimension which might have some bearing on the design of the new one-year course. It was not, however, within the brief of the study to establish a complete picture of the problems obtaining in diverse school contexts throughout Guinea.

1.2. THE PRE- AND IN-SERVICE TEACHER EDUCATION PROJECT

1.2.1. The study was concerned with the impact of the pre-service element of the Pre-service and In-service Education project. The project started in 1985 and terminated in 1993. A project framework, drawn up in 1989, envisaged the following outputs of the project:-

- a. training of sufficient teachers of English to cover all secondary and vocational schools;
- b. the development of an appropriate English Teacher- Training Programme for Guinea;
- c. the establishment of viable English Language Departments at the University of Conakry and ENS;
- d. the emergence of Guinean Teacher-Training Specialists at tertiary level;
- e. the development of support for secondary English teachers;
- f. the establishment of an English syllabus for the 3-year secondary curriculum;
- g. an improvement in the range and quality of English teaching and learning material at secondary level;
- h. the development of adult ELT education; and
- i. the development of an appropriate ELT programme for non-specialist students in four faculties of the University.

1.2.2. Support to the pre-service ELT programme at ENS was therefore only one element of a wider project. A Technical Cooperation Officer (TCO), working full time on this component of the project, left in 1989 and was replaced by one working both on that programme and at the university.

1.2.3. During the three years 1989/90 to 1991/92, total project expenditure amounted to ,1,058,000, including ,547,000 for TCOs and ,347,000 for UK training awards. It is not possible to provide details of project expenditure since inception, nor to isolate expenditure on the ENS Maneah course. Whilst the overall size of the project has been fairly substantial, the input at ENS has been relatively small, with no more than one full time TCO (costing no more than ,50,000 per year). A very rough estimate would be to assume that a third of the total project expenditure was related to the ENS element, say ,350,000 over a three year period.

1.2.4. The 1991 ODA education sector review commended the involvement of local staff from the University and ENS (ISSEM) in this study, thereby contributing towards the achievement of outputs listed in para 1.2.1. (c) and (d). Two quotations are relevant:-

"An important aspect of the research interest of the Department has been the ELT Impact Study."

"The involvement of ISSEM as well as the University of Conakry teachers in the ELT Impact Study will be significant for the future emergence of a cadre of teacher- training specialists."

2. THE STUDY

2.1. INTRODUCTION

2.1.1. Two main types of information were originally collected in the study. Observational data were obtained in order to provide evidence on whether, after training, the experimental group of language teachers faithfully implemented their training on going into language teaching classrooms in Guinea. Test data, on the other hand, were used to establish whether the training had an impact on student language scores over a two-year period.

2.1.2. The observational part of the study sought to establish the extent to which the skills imparted during the training were actually transferred into teaching. It was assumed that if these skills had been well grounded, they might be expected to occur more frequently in trained than in untrained teachers. The study therefore investigated the nature of differences in pedagogical practice between trained and untrained teachers.

2.1.3. It was only later decided that the study might also help provide a certain amount of feedback which could be put to use in developing a one-year teacher-training course leading to a postgraduate Certificate in Education (PGCE). It was intended that this course should eventually replace the current four-year teacher-training course.

2.1.4. The observational part of the study had been designed for the purposes of ex post evaluation for the Overseas Development Administration (ODA), the sponsors of the teacher- training project. It was not originally designed to provide formative data. It was never envisaged that the study would provide an in-depth analysis of the needs of English teachers in the Guinean secondary school system and the problems obtaining in diverse school contexts.

2.1.5. A small-scale tracer study was, however, added to the main study because of the changed role for ISSEM (paras 1.1.5 and 1.1.6 above). This part of the study was to have been the particular responsibility of the Key English Language Teacher (KELT) attached to the Maneah programme, who had drafted a pilot version of a follow-up questionnaire. He was also to take "responsibility for monitoring, and as necessary prompting, the work of the Guinea team" in the intervals between the consultant's visits. The primary focus of this tracer study was to produce a limited amount of formative data which might help in the setting up of a one-year teacher-training programme at the English Section of the ISSEM to replace the original four-year programme.

2.1.6. For this new training programme to be as appropriate and effective as possible, the school experience of the first people to have been put through the existing programme was believed to be of value. Any information that could be tapped from that school experience might be carefully fed into

the design of the one-year postgraduate course.

2.1.7. It is acknowledged that a potentially useful form of data collection is the use of tracer studies of people who have completed a training programme and are back in the target situation. Such studies are not very often carried out. These people are in fact best placed to comment on the training they have received, when they return to the target situation and start working. Tracer study data are often more useful than those collected at the end of programmes. Data collected at the end of programmes are often suspect as participants have not had the opportunity to reflect on, or to try out what they have been taught to do. They are in a better position to do this in evaluations conducted after an intervening period.

2.1.8. On the strength of the above, the tracer enquiry was designed to record detailed reactions from ENS-trained teachers in the Experimental group to the four year programme they had just completed, in the light of their one-year post-training teaching experience. Data were gathered on their retrospective perceptions of the importance of various elements of the course and how well they thought these had been handled.

2.1.9. The test part of the study sought to establish the extent to which skills imparted during the training influenced the pupils' performance. If the skills imparted by such training were well grounded they should occur more frequently in trained than in untrained teachers. The question then addressed was whether more learning benefit occurs in trained teacher classes as compared to untrained teacher classes.

2.1.10. The findings of this part of the study were also intended to contribute to the development of a standard test to go with the textbook EFSA 6^{eme}, but their main purpose was to cater for ex-post evaluation by the donor agency.

2.2. INPUTS AND COSTS

2.2.1. The cost of the Study to ODA included consultancy fees and expenses and four UK training awards. These funded the local project co-ordinator to visit the consultants in 1989 to set up the study, and three local staff to visit in 1992, to complete the data analysis and prepare articles for publication. The cost of the consultancy inputs, travel, and incidental local expenses was ,17,036, and the cost for the TCTP awards (funded from the country aid programme) was ,18,744. The overall UK cost of the study was therefore ,35,780.

2.2.2. In comparison with the Nepal study, which cost ,42,000 in total, all of which was spent on consultants, (see Burton et al 1993), the expenditure on external consultancies for this study was substantially less. Had the KELT attached to the teacher-training project in Guinea remained in place these could have been cut down even further, though the value of the outsider coming in to check on deadlines was useful as a spur to action. The study compensated for the lower consultancy inputs by having the local team mark all the tests and record results on mark sheets, and by having the data entered onto computer disk locally.

2.3. PERSONNEL

2.3.1. The study reported on here was initiated and sponsored by ODA who appointed an external consultant from the University of Reading in the UK to take charge of conducting the initial feasibility

study with stake holders.

2.3.2. A nine member research team was set up, selected from the three Guinean ELT specialist institutions: four teachers from the English Department of the University of Conakry, four from the ISSEM and one from the English Section of the National Pedagogical Institute. In the summer of 1989, the first draft of the evaluation instruments to be used was developed by one of the team members during a short attachment at CALS, University of Reading.

2.3.3. The external consultant paid four short visits to Guinea, first to conduct an initial feasibility study and then to assist in the development of the instruments for use in observation and testing. During these visits he assisted in the training of the team members in observation techniques and test administration. The monitoring of the actual data collection and coding was done at distance by mail.

2.3.4. On completion of the data base, a 10-week attachment at the University of Reading was arranged for three members of the research team, during which they produced a written account of the main results and findings from the baseline study.

2.4. THE SAMPLE

2.4.1. For this evaluation 23 secondary schools were originally selected, covering the 4 main geographical zones of Guinea, as shown in the following table:

Table 1: Schools in the study, according to the geographical zones of Guinea

	GEOGRAPHICAL ZONES				
	LOWER	MIDDLE	UPPER	FOREST	TOTAL
SCHOOLS	13	5	4	1	23

2.4.2. At the beginning of the evaluation two groups of teachers were selected (an experimental group, i.e. trained teachers, n = 14; a control group, i.e. untrained teachers, n = 14). This was an opportunistic sample in that the experimental group was composed of those 14 Maneah graduates out of the thirty who were able to secure teaching posts and the control group teachers constituted the few remaining teachers who had not been trained in UK or Maneah.

2.4.3. The experimental group of trained teachers:

- a. had no background in English at the beginning of the training course;
- b. had successfully completed the four-year training course; and
- c. had no teaching experience at the beginning of the study.

2.4.4. The control group's untrained teachers:

- a. had no formal TEFL training background;
- b. were teaching in places reasonably accessible by road. (The quality of the road was more important than the distance to be covered);
- c. had an average teaching experience of 6 years, ranging from 2-16 years; and

d. had a level of English sufficient to teach the course-books.

2.4.5. Each of the teachers in both groups was allocated one of the 28 classes for a period of two consecutive years. In five schools there were two experimental teachers.

2.5. METHODOLOGY

2.5.1. Three instruments of data collection were used: a Teacher Observation Checklist, a Test Battery, and a Tracer Study Questionnaire.

The Teacher Observation Checklist

2.5.2. Direct information on the classroom behaviour of teachers and students can only be obtained through observation. It must, however, be noted that it is not possible to report everything observed in a classroom. Prior selection of what is relevant and significant is necessary.

2.5.3. To design the instrument an inventory of training characteristics was produced by the trainers for the ENS English teacher-training programme. Then, in discussion between the Guinean research team members and the consultant, 12 criterial indicators of teacher behaviour were selected from the inventory which were considered to be the key defining characteristics of the programme as planned in the syllabus. These key characteristics formed the criterial indicators on which section 3 of the observational instrument was based. These features were meant to represent those elements of the training which the trainers thought were most likely to differentiate between the trained and untrained teacher.

2.5.4. As well as the criterial indicators of teacher behaviour in section 3 of the checklist, the teachers' and students' use of English (as against French) observed during three five-minute periods in each class, were to be coded in section 1 of the observational instrument; a descriptive account of the teaching and learning activities taking place in the whole lesson was recorded in section 2; and an estimate was made at the end of the lesson, in section 4, of the relative time during the whole lesson devoted to teacher and pupil talk, in English and French. It was thought by the trainers that the trained teachers were likely to use less French in the classroom than their untrained colleagues and there was an expectation that there would be a more balanced ratio of teacher/student talking-time in the classes of trained teachers.

2.5.5. The indicators described in the observational instrument used in Guinea thus cover the categories of:

- a. teachers' verbal interaction with pupils in English and French;
- b. the teaching and learning activities taking place in the whole lesson;
- c. presence or absence of certain criterial features in the conduct of lessons; and
- d. the relative time devoted to teacher and pupil talk, in English and French, in the whole lesson.

2.5.6. The observational instrument thus involved four data collection procedures:

- a. coding samples of classroom talk into appropriate categories;
- b. taking unstructured field notes describing the lesson;

- c. completing the checklist of criterial indicators; and
- d. diagramming teacher/pupil talk-time estimates for the whole lesson.

2.5.7. After the initial design of the observational instrument, it was discussed and then tried out by team members. As a result of this process, a number of amendments were made in discussion with the consultant. The consultant also conducted an observer training seminar for the team members responsible for collecting the observational data (for full details of this see Weir, C.J. Report on Guinean Teacher-Training Baseline Research Project : Training Seminar - 3-9 November 1990) and Roberts J. 1990). In the course of this training, videos of classes and live classes in different schools were observed. Observation procedures, and criterial indicators were piloted, reviewed, revised and agreed by the team members and the consultant. At the end of this process, the instrument was considered to be suitable for the collection of data, and inter-observer agreement had also reached an acceptable level.

Conduct of the Observation

2.5.8. Three series of visits were conducted during the academic years November 1990 - May 1992. The first series of visits took place at the beginning of the first academic year, in November-December 1990 when graduates had between 6-8 weeks of teaching experience. Out of the 28 target teachers, 25 were observed (14 from the experimental group, and 11 from the control group) in 21 schools. The three remaining teachers were no longer at their posts when the observers arrived.

2.5.9. The second series of visits scheduled for the end of the first year (April-May 1991) could only be completed at the beginning of the subsequent academic year (November 1991) due to unrest in schools. Out of the 25 teachers expected, 21 were observed in 18 schools (experimental group 12; control group 9). Again a number of teachers had dropped out of the study either for training purposes or as a result of displacement.

2.5.10. The third and last series of visits took place at the end of the second academic year of the study (April-May 1992). Out of the 21 teachers from the preceding visit, 17 were observed (experimental group 11; control group 6) in 16 schools - there had been a further attrition of three teachers from the control group and one from the experimental group. The drop-out in the control group can be explained by the policy of the educational authorities towards improving ELT, particularly in school-leaving examination classes (Baccalaureate). To this end, untrained teachers were likely to be either removed or shifted as soon as a trained teacher was available.

2.5.11. As far as difficulties in implementation were concerned, the observations were often hampered by the non-observance of ground rules by some local authorities as well as severe problems of communication. Moreover, despite careful planning during the sampling, it turned out that some schools were hardly accessible because of the bad state of the roads. To illustrate this, it is worth pointing out that the only Landrover that was available to the team for the visits turned over during one of the trips. This bad state of the roads and the lack of telecommunication facilities explain why it was often not possible to get information to and from school authorities. Add to this an unreliable postal service, and it will be understood that the whole exercise was conducted in very difficult circumstances.

2.5.12. The implications of all this for the study visits were as follows:

- a. teacher observations turned out to be unpredictable e.g. the teacher was not present

for social reasons or because of recently shifting to another locality, or the teacher was there but was sick, had been assigned a non-experimental class, or had shifted to a different school across the road. In short, it was only on the spot that this information became available, making it very difficult to make alternative arrangements; and

b. by the end of the series of visits some 11 teachers were out of the study.

2.5.13. Only those teachers who had been observed during the first visit and had kept the same class were observed during the second visit. On the third visit only those who were observed at the second visit and had kept the same class throughout the study were observed. They were a total of 17 in number and constitute the core of the forthcoming analysis.

2.5.14. Reliability was enhanced through some observations being carried out by two independent observers. This pair of observers was different at each visit and for each teacher. At the end of each observation an agreed copy of the observation schedule was produced by each pair.

The Test Battery

2.5.15. The tests were based on the textbook in use in secondary schools: "English for French-Speaking Africa". It was assumed that, on the basis of 3 hours teaching per week (maximum 22 weeks), only the first 15 units of the book would be covered during the school year. So the test battery was devised to test only those 15 units in its administration at the beginning and end of the first year. The implementation of the study required the first test to be carried out at the beginning of the first year of the study, the second at the end of the same year and the third and final test at the end of the second year of the study. Further elements were added to the test at the third administration to cater for the learning that had taken place in the second year of the study.

2.5.16. The battery had four components: Reading (gap-filling tasks), Writing (direct writing tasks), Listening (dictation task), and Oral (direct speaking task); and included sentence repetition, interview, and information transfer sub-tests. It was felt that estimates of learner intake would shed light on teacher performance, and that pupils' performance on the test battery would constitute the best indicator of learner intake.

2.5.17. For the first and second administrations of the test, the battery used was based on the first 15 units and included the following number of items per constituent:

Reading Gap-filling - 70 items,

Writing - 20 items,

Dictation - 43 items;

and a speaking test which had three elements:

Imitation,

Information transfer,

Interview.

2.5.18. The third test included some items from the last 15 units of the book in addition to those from the first 15 units.

Table 2: Number of Items in Tests Administered

Test Component	Number of Items		
	Visit: 1	Visit: 2	Visit: 3
Gap-filling	1-34	1-34	-
	35-70	35-70	1-36
	-	-	37-75
Writing	1-20	1-20	1-20
	-	-	21-35
Dictation	1-43	1-43	1-43
	-	-	44-53

2.5.19. This table shows an increase in the number of items in each element of the test in the third administration except for Speaking which remained unchanged throughout. So, finally, there were -

Reading(Gap-filling) : 75 items including 39 from the last 15 units.

Writing: 35 items.

Dictation: 53 items.

2.5.20. Students from both groups sat the same tests, at the beginning and end of the first year of treatment. A supplemented version of the test was then administered at the end of the second year of the study. It should be noted that as far as the oral component of the battery is concerned, practical reasons dictated that only a sample from those who took the written part were tested.

Piloting

2.5.21. The pilot version of the test battery having been designed, it was first discussed and then tried out by team members. Classes of pupils were also organised in a local school to enable the team to carry out some dummy oral tests so that they could iron out any difficulties in this part of the study.

2.5.22. In response to the piloting some alterations were also made in the scoring system of such test elements as Guided Writing and Dictation. Double marking was agreed for all parts of the test battery to enhance reliability (For details see Weir, C.J. Report on Guinean Teacher-Training Baseline Research Project : Training Seminar - 3-9 November 1990).

Conduct of the Test

2.5.23. Three series of visits were conducted during the period November 1990 - May 1992.

2.5.24. The first series of visits took place at the beginning of the academic year, in November-December 1990. During these visits pupils were tested on the four components of the test battery in 21 schools, 25 classes held by 25 teachers (Control Group teachers:11; Experimental Group teachers: 14).

2.5.25. The second series of visits scheduled for the end of the first year (April-May 1990) could only be completed at the beginning of the subsequent academic year due to unrest in schools. During these visits pupils were tested on the 4 components of the test battery in 18 schools, 22 classes held by 22 teachers (control group: 9; experimental: 13).

2.5.26. The third and last series of visits took place at the end of the second academic year of the study (April-May 1992). Out of the 22 classes from the preceding visit, 18 were tested (control group: 6; experimental group 12) in 16 schools. There had been a further attrition of four classes (control group: 3; experimental: 1).

2.5.27. This drop-out can be partly explained by the policy of the educational authorities towards improving ELT, particularly by employing trained teachers for school-leaving examination classes (Baccalaureate), and partly by pupils changing school as a result of parents' shifting place of residence.

2.5.28. In general, untrained teachers were likely to be removed from teaching as soon as a trained teacher was available but in some cases attrition was due to families moving to live in a remote part of the same town (or in a different town).

2.5.29. The difficulties which affected the conduct of the teacher observation checklist (see para 2.5.11 above) applied equally to the conduct of the test battery. The implications for the study visits were as follows:

a. Reduction of the number of classes:

Visit 1: control group: 11; experimental group: 14

Visit 2: control group: 9; experimental group: 13

Visit 3: control group: 6; experimental group: 12

b. Reduction of the number of pupils:

By the end of the study the following data were available:

Table 3: Numbers of Candidates Providing Test Data

Visits	Test element	Number of Candidates		Total
		Control	Experimental	
	Gap-filling	281	413	694
1	Dictation	288	432	720
	Writing	289	425	714
	Gap-filling	204	280	484
2	Dictation	204	286	490
	Writing	201	290	491

	Gap-filling	115	166	281
3	Dictation	98	142	240
	Writing	116	172	288
1		58	72	130
2	Oral	35	43	78
3		16	21	37

2.5.30. This table indicates severe attrition from the first through the third visit i.e. as far as the written part of the test is concerned, the study started off with a mean of 709 pupils. This came down to 488 pupils during the second visit, and finally reached the number of 269 pupils during the third visit (some of these students, however, were lost because of faulty numbering of certain candidates by the team members).

2.5.31. The oral part of the test was affected by the same attrition. The number of pupils who took part in the oral test dropped from 130 in the first visit to 78 in the second, and finally to 37 during the last visit.

The Tracer Study Questionnaire

2.5.32. This included a rating of the different training course components and the making of open-ended comments on difficulties encountered by experimental teachers in their teaching contexts. It was reasoned that this would provide some feedback on the relevance of the course components to individual teaching situations as well as on the amount of training time devoted to each component.

Design

2.5.33. A questionnaire with two main sections was developed by the team members and the KELT attached to the Maneah project. In the first section the graduate teacher is asked to make judgements of sufficiency and importance on the 7 main elements (i.e Language Development, Literature, Methodology, In-House Practical work, Teaching Practice, Other Components, and Written Assignments) of the original four year Maneah training programme. It was made clear that teachers' choices should be based on their post-qualification experience.

2.5.34. In the second section, graduate teachers were asked open-ended questions of the type: What do you think of....; Do you have comments about....?, and to make comments in response to prompts such as: Give your assessment of....; Give comments on....; Identify....; (see Table A7a, A7b, p47,48).

2.5.35. Due to time constraints and, according to team members, the non-availability of respondents, a trial stage was not carried out before the first version of the questionnaire was administered.

2.5.36. The first administration of the questionnaire in the event turned out to be the piloting phase. It was realised at that point that the instrument needed some improvements. As a result of the first administration of the questionnaire, the first part remained unchanged while some alterations were made to the second part, eg., the lay-out was altered as well as the wording of questions 1 and 5, and

more space was provided for comments.

Administration of Tracer Study Questionnaire

2.5.37. Tracer study data were obtained by means of a final questionnaire distributed to the 12 experimental group teachers remaining, at the outset of the second year of this study. The questionnaire was kept by the target teachers until the end of their second year of teaching and was handed in or sent to the observer team when completed to their own satisfaction. The distribution was carried out by the observer team during a second series of observation visits. The graduates were asked to go through the questionnaire and clarifications, when needed, were given by the team before leaving. It was emphasized that the study aimed at evaluating the training course itself, not the teachers. To what extent this was understood, and to what extent questionnaire responses were distorted by such matters as the opinions and feelings of the individual, is not evident. With this form of enquiry it is inevitable that the object of the enquiry, i.e. information about training, will be inseparable from the effects of various individual factors.

2.6. PROBLEMS DURING IMPLEMENTATION

2.6.1. There were some problems during the study arising out of the difficulty of communicating between UK and Guinea Conakry. With no telecommunications system and an erratic postal service it was not always possible to maintain the degree of contact one would have liked. The departure of the KELT after the first year of the study also meant that there was very little external control on the manner in which the research was carried out. A higher degree of monitoring than is possible from distance would have been useful, particularly during the collation of data and also during the data collection trips.

2.6.2. Counterparts need to gain the confidence to do things for themselves but this must be preceded by a phased initiation into research skills. This is best done by the KELT responsible for the project who is in daily contact on the ground with other members of a research team. If this is not possible then some reduction in professional rigour may be inevitable.

2.6.3. The severe social disruption and economic hardships present throughout the study certainly affected the motivation of a small number of the research team. This led to some difficulty in getting the data marked and collated and resulted in a certain lack of rigour in parts of the research exercise. This is not to detract from the commitment of people who participated fully until the very end and contributed substantially to its successful conclusion.

2.6.4. The final attachment of the three Guineans to Reading to write up the study in a number of papers was a valuable experience for all concerned. It was at the collation and analysis of data stage that the looseness with which some of the data had been collected, recorded and entered onto the computer, really came home to the three Guineans. The problems this caused at the writing up stage was a salutary lesson. In the discussion which led to the sorting out of these problems they finally began to take on board the need for a rigorously systematic approach in research exercises of this type. The involvement of counterparts in such studies provides a valuable learning experience in research methodology. Having counterparts write up the results of a study for possible publication also provides them with an extremely motivating enriching experience. It has made an important contribution to the potential for sustainability of the project and to the future professional self-respect and commitment of the three involved.

2.7. RESULTS

Results of the Observations

2.7.1. By the end of the study, the following data were available for analysis:

- a. 17 teachers had been observed on all three visits, 11 from the experimental group and 6 from the control group;
- b. a total of 51 observations were carried out in 16 schools (2 teachers from the experimental group worked in the same school); and
- c. these observations were supplemented by tracer study questionnaires (for details, see Annexes 3 and 6).

Use of Language

2.7.2. From the observational data, the percentages of the use of English or French by both teacher and pupils have been estimated for the individual classes in the Control and Experimental groups (see Table 4). These percentages are based on the mean percentages of Teacher/Pupil (T/P) language over the three five-minute segments coded for each class.

Table 4: Mean Percentages of Teacher/Pupil Language Use Per Series of Visits

Codes	Teacher						Pupil					
	E1	F1	E2	F2	E3	F3	E1	F1	E2	F2	E3	F3
UT1	72.0	4.5	69.9	0.0	61.7	1.4	19.3	4.1	28.2	1.8	36.1	7.0
UT2	76.2	0.0	56.3	0.0	54.4	4.4	22.4	1.4	43.6	0.0	40.0	1.1
UT3	67.9	12.0	69.7	0.0	63.2	0.0	19.1	0.9	30.3	0.0	36.7	0.0
UT4	57.9	3.1	56.0	6.6	60.0	4.5	39.0	0.0	35.5	1.8	35.5	0.0
UT7	58.9	1.6	85.8	0.0	82.4	0.0	39.0	0.2	14.2	0.0	17.5	0.0
UT8	60.7	10.7	58.8	2.0	53.4	2.7	28.5	0.0	39.1	0.0	37.2	6.7
Mean of Means												
E	65.5		66.0		62.5		27.9		31.8		33.8	
F		5.3		1.4		2.1		1.1		0.6		2.4
T1.	73.9	0.0	66.2	1.5	60.7	2.0	26.0	0.0	31.1	1.0	36.2	1.0
T2.	69.1	9.1	82.6	0.3	72.2	0.4	20.6	1.0	16.7	0.3	27.4	0.4
T3.	56.2	0.7	71.0	0.6	80.4	0.4	42.3	0.7	27.0	1.2	19.5	0.0

T4.	54.5	5.4	65.4	0.0	72.4	0.3	37.0	2.9	34.5	0.0	26.9	0.3
T6.	47.0	0.0	63.9	1.9	55.7	0.0	52.0	0.0	31.2	0.8	44.3	0.0
T8.	65.7	0.3	66.1	1.5	49.8	0.5	31.8	1.9	32.3	0.0	45.5	4.0
T9.	51.7	0.6	69.4	0.5	71.9	0.0	47.2	0.3	27.1	2.8	28.0	0.0
T10.	53.2	19.8	60.9	7.1	51.8	2.2	19.1	4.2	27.8	4.0	25.9	0.0
T11.	57.8	1.8	65.4	0.3	75.0	0.0	39.8	0.4	29.3	4.7	21.4	2.6
T12.	57.7	0.0	67.6	3.8	80.0	0.0	40.3	0.0	25.1	3.4	20.0	0.0
T13.	66.3	5.6	75.6	0.8	71.4	0.0	28.0	0.0	23.4	0.0	28.6	0.0
Mean of Means												
E	59.3		68.5		67.4		34.9		27.7		29.4	
F		3.9		1.6		0.5		1.0		1.8		0.8

Horizontally, Letters refer to the language used : E = English, F = French, and Numbers refer to the series of observation visits: 1 = first visit, 2 = second visit, and 3 = third visit.

Vertically, UT refers to control group teachers and stands for Untrained Teacher; T refers to experimental group teachers and stands for Trained teacher.

2.7.3. The table seems to indicate that:

- a. despite a generally consistent picture of co-variation of Teacher Talking Time (TTT) with Pupil Talking Time (PTT) there is no consistent pattern from teacher to teacher in the changes in TTT which take place over the three visits. One teacher might have a high TTT on the first visit, and a low one on the second, while another might have a low TTT on the first visit, and a high TTT on the second. (See UT7, and T6 - T13, Table 4, page 20);
- b. all classes observed tended to be teacher-centred;
- c. English was the language most used in all classes observed, irrespective of whether groups were control or experimental; and
- d. apart from these general trends as shown in the table above, no significant differences were found between the two groups, with respect to either teacher or pupil. Thus, there does not appear to be any marked difference between experimental and control groups in the overall mean percentages of language use by pupils in English (E) and in French (F), or by teachers in English and in French, from a quantitative point of view. This is more clearly illustrated in the following summary tables.

Table 5: Overall Mean Percentages of Pupil/Teacher Talking Time over the Three Five - Minute

Segments Coded for Each Class and for the Three Visits

	TEACHER		PUPIL	
	E	F	E	F
CONTROL	64.7%	2.9%	31.2%	1.0%
EXPERIMENTAL	65.1%	2.7%	30.7%	1.2%

Table 6: Overall Mean Percentages of Estimates of Pupil/Teacher Talking Time Over the Whole Lesson for Each Class and for the Three Visits

	TEACHER		PUPIL		
	E	F	E	F	ZERO
Control	52.4%	4.2%	30.9%	2.1%	10.2%
Experimental	54.0%	2.3%	34.6%	1.6%	7.0%

NB. ZERO = stretch of time when no talking is involved

Checklist Data

2.7.4. During the three visits data was collected on teacher classroom behaviour using the checklist of criterial indicators. It appears from these results that the "checklist" data, summarising the observed indicators, shows a significant difference between the teachers in the Experimental group and those in the Control group in all the indicators except one, as shown in Table 7 below.

Table 7: Mean Totals of Occurrences of Criterial Indicators 1-12 Over the 3 Visits, and Mean Occurrences Per Teacher, for the Experimental and Control Groups

Criterial Indicators												
	1	2	3	4	5	6	7	8	9	10	11	12
Experimental Group (N=11)												
Total Occurrences	13	13	14	9.7	6	5	1.1	2	7.3	3	14	12
Occurrences per teacher	1.2	1.2	1.3	0.9	0.5	0.5	0.2	0.2	0.7	0.3	1.3	1.1
Control Group (N=11)												
Total Occurrences	5.7	6.3	5.3	1.0	2.6	2.0	0.0	1.3	2.0	1.3	5.7	3.6

Occurrences per teacher	0.9	1.1	0.9	0.2	0.4	0.3	0.0	0.2	0.3	0.2	0.9	0.6
-------------------------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

2.7.5. It will be seen that the experimental group teachers show a higher occurrence of all indicators except 8 (Pupils did guided writing not in book). Given that control group teachers exhibited a similarly low frequency on this activity, it may be that the practice of guided writing needs to be more thoroughly grounded in the training programme. Alternatively it might well be that this is an activity which is not considered to be appropriate in the current pedagogical context of the Guinean secondary schools. Teachers' views on their needs for further training, as reported in the tracer study, support the latter view (see below). Low incidence of this feature in either group may well attest to its lesser importance.

2.7.6. There is a complete absence of indicator 7 (Teacher gave extended listening practice) in the control group, and in the experimental group it hardly occurs either. It can be argued that, unlike indicator 8, the non-occurrence of indicator 7 suggests that more emphasis has to be laid on this in the forthcoming one-year course. In fact, listening practice appears to be reported as one of the main difficulties for the Maneah graduates in the tracer study conducted as part of the baseline study.

2.7.7. There is, however, a crucial issue behind the implementation of activities 7 and 8 which relates to the pedagogical context obtaining in Guinea. This concerns the issue of transfer of skill from training site into the language classroom, in as much as, the newly trained teachers, in their attempts to influence the curriculum, are limited and sometimes influenced by, the school situation. The teaching one can see happening is simply the result of this sum of interactions between the teacher and the school setting. It points once again to the need for appraisal studies of the contextual situation in school systems to take place before a syllabus for a teacher- training programme is ever drawn up.

2.7.8. The occurrence of the following indicators in the experimental group teachers is more than 30% higher than for the control group:-

- 1 teacher presented language in situation, mainly in English;
- 3 oral drills used in English - more than one student;
- 4 pair-work/Group-work used i.e all the class involved;
- 11 teacher showed control and management of his class;
- 12 the aims of the lesson were clearly defined and remained so.

It is here, in these basic pedagogical techniques, more than anywhere else, that training is seen to make a significant difference, despite the gap in the number of years of experience between experimental and control group teachers. In respect of these criterial features the programme can be seen to have benefitted the experimental group teachers.

2.7.9. On indicator 2 there is a relatively high, and very similar frequency of occurrence in both groups (Teacher gave model in English i.e. word/sentence). There is also a high frequency of occurrence in respect of indicators 1, 3 and 11 in the control group (though lower than for the experimental group); this attests to the fact that good experienced teachers will naturally, in the course of time, adopt certain of these criterial features in their pedagogy. Indicators 4 and 12, however, indicate that experience in itself will not supply teachers with all desirable techniques and practices.

2.7.10. The indicators occurring least frequently in both groups appear to be:-

- 5 comprehension questions asked in English i.e - new ones not in book;
- 6 pupil produced original sentence in English - at least 3 different pupils;
- 7 teacher gave extended listening practice;
- 8 pupils did guided writing not in book; and
- 10 extra practice - e.g own drawings/questions.

These areas are perhaps the most communicatively demanding for both the teacher and pupils, and they are obviously in need of greater attention in the new teacher-training programme if they are to feature more strongly in the Guinean classroom. Given that they were considered to be important distinguishing features of the trained teacher at the start of the study, their absence in implementation is serious and calls for further investigation. The designers of the new programme need to establish whether their absence is due to contextual constraints arising out of the school situation, as suggested above, or because these features were insufficiently well-grounded in the original training programme.

2.7.11. The data from Sections 1 and 4 of the observations also suggest that the trained group did not vary significantly from the untrained group in the amount of teacher (as against pupil) talking time in class. The expectation of the trainers was that there would at least be a balance in the trained teachers' classes and that an ideal result would be that the pupils would be given the opportunity to talk more than the teachers.

Test Results

2.7.12. The final figures for students involved in the analysis of results are presented in the Annex Tables.

Discussion of Results

2.7.13. Only 8 completed questionnaires from the group of 14 experimental teachers were received and this obviously represents a restricted sample of the current teaching force who have been trained solely in Guinea. The issue of sampling is as crucial here as it is in all forms of educational enquiry. Figures are quoted for the 8 teachers sampled in the tables above, but they need to be treated with caution, given the small size of the sample.

2.7.14. The majority of respondents agree that all the 7 elements of the course are, if not very important, then important. With the possible exception of non-specialist teaching, very few think that they are not important. On the issue of whether or not training was sufficient - i.e in terms of time allotted and learning benefit - over one-third feel that enough attention was given to each element. About one-third of the respondents, however, are dissatisfied in this regard and almost a third are undecided. Although there is a general agreement on the importance of the training elements, quite a sizeable proportion of respondents think that insufficient attention was paid to nearly all the pedagogically-oriented course components.

2.7.15. The tracer study questionnaires revealed a breakdown of trainees' perceptions of the sufficiency of individual course components at Maneah:

- a. in Language Development the subjects that appear to need most attention are Listening Comprehension, Integrated Communicative Activities and, to a lesser extent,

Formal Grammar and Speaking;

b. in Methodology: there is evidence of need for some improvement in Applied Linguistics, the Teaching of Reading, Speaking, and Classroom Management;

c. as far as In - House Practical Work is concerned greater emphasis should be placed on Discussion & Use of EFSA and Non-Specialist Teaching;

d. in the Teaching Practice component there is a clear need for increased stress on Assistance from Tutors;

e. concerning Other Components, Outside Speaker and School Administration Lectures appear to need greater attention; and

f. written Assignments in Literature require more careful planning and supervision.

2.7.16. Based on the data available, replies to the open ended-questions indicate that:

Question 1.

The balance between the two major strands of the course, i.e. language development and professional training, is not quite right. Indeed, on the question of balance between the two, 4/8 of respondents don't know and 1/8 thinks there is not a good balance, whereas only 3/8 agree that the balance was right. This high percentage of "don't know" respondents could be explained either by the fact that the question was not explicit enough despite changes made to the questionnaire's original version or that they were uncertain as to whether the balance was satisfactory or not.

Question 2.

Good use appears to have been made of teaching materials in general but the use of listening materials and audio-visual aids could be improved. From comments made on resources available during training:

- 6/8 think that the reading materials were useful (if not very useful), and 2/8 don't know.

- 3/8 think that the audio listening materials were useful (if not very useful), 1/8 think they were less useful, and 4/8 don't know.

- 3/8 think that the video materials were useful (if not very useful), 2/8 think that they were less useful and 3/8 don't know.

Question 3.

The ENS methods of teaching appear to be appropriate. From comments made on these methods 100 % of respondents think that they were good, if not very good (this positive reaction to the teaching was also evident in a UK Project Officer's earlier reports).

Question 4.

A seminar in an English-speaking environment appears to be of value. Indeed, 100 % of respondents think that the seminar in Bo (Republic of Sierra Leone) was very useful or useful.

Question 5.

A certain number of training areas still cause difficulties to 5/8 of respondents. The areas identified were:

- the use of French in ELT
- management of a large class
- coping with lack of teaching materials
- teaching receptive skills (listening & reading)
- teaching productive skills (oral)
- teaching phonology (pronunciation)
- teaching grammar
- evaluation & testing (assessment)

3/8 respondents don't know, however, and few of the respondents offered five suggestions. This percentage of "don't know" respondents could be explained by the fact that the question was not explicit. It also highlights the type of problem which derives from the use of an open-ended format in enquiries of this kind and also from the nature of questionnaire surveys in general. The data might have been more complete if the questionnaire had been used in a semi-structured interview mode as had originally been intended. The interviewers would then have been able to prompt, clarify and elicit responses with individual respondents. This might also have reduced the numbers in the "don't know" category.

Question 6

4/8 respondents feel that they need further training.

4/8 respondents answered "don't know" to the question and did not make any additional comments.

2.8. MAIN FINDINGS OF THE STUDY

Conclusions from the Observational Data

2.8.1. It was noted in the first consultancy visit (see Weir 1989:2):

" The methodological components of the Maneah course are those most in need of serious attention. It is clear from discussions with tutors on the course and with a UK Project Officer that it is only in the last year they have seen themselves as a teacher-training department, the majority of the earlier impetus having gone into language training (the success of the latter can be

judged from the Teacher's results). The upshot of this is that the methodology components and the teaching practice were not as well implemented as the UK project officer would have liked..."

2.8.2. Despite the earlier problems in implementation, however, the observational data on teachers showed a number of marked differences between the classroom practice of the experimental group teachers and the control group teachers. The reliability of the findings was enhanced by the use of paired observations in every case.

2.8.3. In terms of accountability, the observation procedures described in this paper can supply data which shed light on whether certain identified key features of training have been implemented. They will provide a valuable formative tool for monitoring the teaching practice component of the future one year preservice teacher-training programmes in Guinea. They will provide trainers with a systematic instrument for monitoring the take-up of training and could assist trainees in their own cooperative monitoring of their peers.

2.8.4. In spite of the long experience of teachers in the Control group, the differences in criterial indicators between the two groups, in favour of the experimental group, may well prove important. They might have a significant impact on classroom learning experiences for pupils in the long term.

2.8.5. The formative value of these data from the observations for the design of the future teacher-training programme is obviously restricted. The observational procedure of this study was constrained by the variables it was designed to identify. Its primary purpose was to identify the presence or absence of pedagogical features deemed by the stake-holders, the trainers themselves, to have been important elements of the first training programme. It does not take account of any variables in the teaching context other than those identified. The light it sheds on the identified variables does of course provide valuable food for thought.

2.8.6. For accountability purposes the closed type of data generation instrument discussed above is useful. To generate more illuminative data which might be of value in the design of a new programme or the revision of an old one, more open ended data collection methods would need to be employed. The closed type of observation instrument employed in the study was designed primarily for summative purposes. An observation instrument with open categories would be more suitable at the appraisal stage in project design when data are normally collected to help design a programme which is appropriate to the teaching context obtaining in a particular educational milieu.

2.8.7. The formative dimension was more directly addressed in the tracer study part of the research, in an attempt to elicit past students' reactions to the training programme in the light of direct teaching experience and to try to identify areas which they felt should receive more attention in the future training programme.

Conclusions from the Test Data

2.8.8. The results above are the raw (unadjusted) scores for each exam at each visit for the control and experimental groups separately. In general the results indicate significant improvement from visit 1 to visit 2 but smaller changes from visit 2 to visit 3, only significant for gap-filling in both groups and for writing in the experimental group.

2.8.9. Further analysis revealed no significant differences between the scores of the control and experimental groups in any of the tests. The other variables of numbers in class, number of books and language ability of the teachers did not have any significant effect. On the whole there was a non-significant but slightly better improvement in performance on the part of the experimental group

by the end of the second year of the study when the results are compared with those obtaining at the end of year one (visit 3 - visit 2 in the Tables above). Over the two year period (visit 3 - visit 1) the experimental group showed more improvement than the control group in the dictation and gap-filling tests but not in the writing test, although none of these differences in improvement figures between the groups were significant.

2.8.10. The data suggest no significant differences between the scores of the two groups at the start of the study and no significant differences between the groups in terms of language improvement over the two years of the study. This lack of any significant difference between the performance of the two groups at the end of the two year period of the study can be interpreted in a number of ways. The positive interpretation is that the performance of the students in the classes of trained teachers, with no previous classroom experience, is already comparable with that of students taught by experienced teachers. The negative interpretation is that training has only made a slight but not a significant difference in terms of the impact on student language scores.

2.8.11. The raw scores of the experimental group do, however, suggest a slight trend in favour of this group. This is perhaps an indication that the impact of the training is beginning to appear and that it would be useful to conduct such baseline studies over a longer period of time for the full effect of such training to become manifest in student test scores. The attrition of groups in studies such as this (see Para 2.5.30 and 2.5.31) may, however, pose problems for an extended study. The possible need for a longer study period and a larger initial sample has obvious implications in terms of cost and logistics.

2.8.12. Only a very rough estimate of the number of hours taught in these two years was available, because of the complete lack of any record keeping in the schools despite repeated requests for this from the start of the study. Given the limited number of hours theoretically available for teaching English (120 hours in 1990-1992) and the reduction of this due to civil strife and other problems (c100 hours), the small gains in student scores for both groups appear reasonable. There appears to have been quite a serious drop in improvement in the second year of the study (visit 3 - visit 2 as against visit 2 - visit 1) which might be attributed to the chaos caused by civil strife.

Conclusions from Tracer Study Data

2.8.13. The course appears on the whole to have been well designed with most elements felt to have been important. There is some indication, however, that the methodological and practical components may not have been dealt with sufficiently. To a certain extent, the integrated nature of the course, involving a considerable amount of language instruction throughout, accounted for this state of affairs. In any event, a PGCE course would be enhanced through the allotment of more time and attention to training in methodology, teaching practice and in-house practical work.

2.8.14. As far as teaching materials are concerned those for reading appeared to be more useful than others. The listening and audio-visual materials would appear to require more attention. The Maneah teaching method's actual delivery appears, on the whole, to have been well received but it will obviously need adaptation to suit the PGCE course's more restricted timespan.

2.8.15. For those trainers involved in carrying out the study, the study has highlighted the need for rigour and a systematic approach. Reliance upon the goodwill of the trainees to return the questionnaires, and upon an inefficient postal system to facilitate this was obviously naive. It was unfortunate that the original plan to use the questionnaire in a semi-structured interview with the Maneah graduates was not followed. Nor can the need to test the suitability of questionnaires before they are used be stressed too strongly.

2.8.16. Questions still linger over the validity of sampling, and the general applicability of results. Fifty seven graduates from the first and second cohorts from the Maneah programme are now in the school system. It is intended that a revised questionnaire be administered to the first and second cohorts from the programme when they attend the next national seminar in Conakry. The Maneah programme graduates will be required to attend a special session where the purpose of the revised questionnaire will be explained to them and help will be available to assist them in completing it forthwith.

2.8.17. In addition, the future graduates from the new Maneah programme will be followed up in a similar fashion to ensure that attention is paid to improving the new one year programme on a continuing basis. In this way a commitment to educational improvement and understanding, as well as accountability, will be built into the system. The value of such an illuminative approach will be clear to all trained teachers and will, one hopes, encourage them to reflect also on their own pedagogical practice. Improvement in any educational system can only come through a commitment to evaluation and a willingness to act on its results.

2.8.18. A tracer study of the type outlined in this paper affords an informative contribution to evaluation; it is both illuminative, and can serve to satisfy the demands of accountability.

3. THE USE OF BASELINE STUDIES IN THE MONITORING AND EVALUATION OF EDUCATION PROJECTS

3.1. A purpose of this evaluation was to provide lessons for those considering similar studies for future projects, especially in the light of the findings of the earlier Nepal study. This chapter of the report attempts to provide some reflection on the efficacy of the baseline study in the context of the Guinea project and to relate the findings to the wider issue of the value of such studies in general.

3.2. As a baseline design has controls for diverse entry levels it is the most appropriate means with which to show the effect of different inputs in terms of gains from the baseline point.

3.3. A baseline design may be most relevant where a project is ongoing because it can then inform subsequent project phases and the costs of such studies relative to the overall project cost would then be lower. They may also be justified in pilot projects which are, subsequently, likely to be replicated or expanded. Baseline studies need to be long-term and for this reason are better suited to longer and larger projects.

3.4. The severe socio-political instability and civil disruption that Guinea was experiencing throughout the study, inevitably had a profound effect on professional morale and commitment. The knowledge that the project was soon to be terminated must also have made it difficult for some local staff to maintain a high level of enthusiasm. As a result of these factors, only about half the team appeared to be involved and contributing to the study. There were problems in getting the staff to collate and mark the data. A certain amount of carelessness in data collection, entering the results, labelling the data and administering the questionnaires was, in retrospect, perhaps unavoidable given the context in which the staff worked.

3.5. One could, however, view this positively as a valuable learning experience for those 4-5 team members who were genuinely concerned with the study's outcome. Certainly the staff who had to deal with the problems of this earlier carelessness were very much aware of the difficulties which this led to at the stage where data was analysed.

3.6. The results of a baseline design are most meaningful if the measured gains can be assessed against a benchmark set by local needs. In this way a benchmark language proficiency target could be set for each situation, for example for entry to further study or to meet employment demands, and would enable a pragmatic assessment to be made of the merit of summative scores.

3.7. Future baseline studies should attempt to define a standard of education required for the labour market and educational systems and monitor student performance against that standard. The cost of such an exercise should not be underestimated. Performance of students could be monitored against these benchmarks.

3.8. An analysis of use of English in the labour market should in any case precede any substantial ELT project and baseline study. Unless English has importance in the labour market even a highly successful project can have no economic benefit.

3.9. As already noted, an adequate externally-conducted baseline study would be most relevant to a large project, where the commitment of resources it demands would not be disproportionate to the total budget. Collecting and processing data in a baseline design that is wholly conducted by "outsiders" requires considerable resources. If local players are involved in the study, the possibility that bias and personal interest may affect the data needs to be borne in mind. The neutrality of the outside, professional consultancy guards against this, and there will always be a need for some external moderation to lend credence to the results.

3.10. In the case of small projects, insider staff could be primarily responsible for collecting baseline data, according to an externally agreed and monitored plan or framework. The evaluators would argue for the involvement of insider staff in baseline data collection in all projects and programmes but would maintain there is a need for some external monitoring.

3.11. There is also, however, a formative dimension to insider involvement, not least in the potential gain in research experience for the staff conducting the study. Some form of continuous monitoring by professionally qualified project staff, greater than was available in the Guinea study, is necessary to counteract this.

3.12. In the case of aid projects, it might be more cost effective to employ local consultants as external evaluators, under TCO supervision; unlike Nepal, this was never a viable option in Guinea.

3.13. Given the likelihood of attrition, sample sizes should be large enough at the outset to sustain the study through its length. The longer the period of the study, the larger the sample groups should be. If, however, there is only a single small trained team and it takes at least a day to visit a school, (and longer where the transport infrastructure is undeveloped) the use of larger groups will mean that it takes longer to complete a series of visits. It thus becomes more likely that the pupils in the last school will have had a much greater exposure to English lessons than those in the first school visited. A bigger team would address this difficulty but would bring with it its own problems. As it stands, the Guinea sample is very small and, as in Nepal, the problem of the general applicability of the results must remain.

3.14. The use of language test data to monitor gain in student language achievement is supported. Their use can provide clear lines of development for the testing of language at all levels within a school system. The design process underlying the construction of valid, reliable and practical tests also requires project staff to be very clear about what it is that the project should achieve. In the long term, such tests could offer illumination in those situations where the existing school leaving examinations are at odds with the objectives of the intended training. In countries such as Guinea,

where no tests are supplied with the school textbooks, the production of such tests can be a valuable spin off from such baseline studies.

3.15. Language tests have to be selected with great attention to their appropriateness and their relevance to the programme objectives. Given the degree of unpredictability in the development of a programme it is prudent to employ a range of tests.

3.16. The observation checklist, the observer training manual and the use of self-report data, offer a tried methodology for the monitoring of future teacher-training projects; they also have a critically important contribution to make to the clarification of objectives during the design of projects.

3.17. It is necessary to obtain data upon the realities of implementation and not to be content solely with test data. This is because an evaluation should be capable of showing a relationship between gain scores and implementation patterns. If there is no difference between the two groups (control and experimental) in terms of the test data it is important to establish the reasons. In such a case the observational data can help establish whether there was a problem with the training and in what areas this manifested itself, ie it helps illuminate the effectiveness of the training programme. Additionally, as in the case of Guinea, there may well have been differences in terms of implementation but no visible effect on student performance in the short term (c.100 hours).

3.18. Furthermore, implementation data have a formative value within the programme, for subsequent programmes, and for a wider audience. It is important that teacher-training projects should be able to provide a clear picture of what measurable changes in teachers' classroom performance they intend to achieve. There should be, from the outset, a clear definition of the objectives of training. This is best served by establishing, with the training and its context in mind, a trained teacher profile which should include a criterial list of features to distinguish between trained and untrained teachers.

3.19. Direct observation, however, is demanding of resources and skilled personnel and lower cost forms of monitoring might be required (eg limited direct observation data cross-referred to teacher and student self-report data).

3.20. Baseline studies are more likely to be acceptable to the recipient country if other spin-off benefits for them, as well as for donors, can be identified.

APPENDIX 1: STUDY DOCUMENTATION

Weir C.J., and Roberts J. 1993

The Role and Design of Baseline Studies in the Evaluation of English Language Teacher Training: the case of Nepal.

Teaching Nepal ODA Evaluation Report Ev 485, Overseas Development Administration: London

Michie E. 1987

Evaluation of ENS English Course (A KELT Project), Maneah, Guinea

Roberts J. 1990

Guidelines for the Analysis of Observation Forms (Guinea).

Weir, C.J. 1989

Report on a First Visit to Assess the Feasibility of a Baseline Study in Guinea Conakry, June 6-11 1989

Weir, C.J 1990

Report on Guinean Teacher-Training Baseline Research Project: Training Seminar - 3-9 November 1990

Weir, C.J. 1992a

Guinean Teacher-Training Project Report on Third Visit - 4-11 January 1992

Weir, C.J. 1992b

Guinean Teacher-Training Project: Report on Fourth Visit June 15th-19th 1992

Weir, C.J. 1992c

Collecting Language Test Data-Final Phase 1992

APPENDIX 2 CHRONOLOGY OF STUDY

June 1989 First Visit by Dr C.J. Weir (CJW) to Assess the Feasibility of a Baseline Study in Guinea Conakry, June 6-11 1989

July 1989 Attachment of Abdul Bah, Guinean research coordinator, to CALS to design draft observation and test instruments

November 1990 Second visit by CJW to conduct Teacher-Training Baseline Research Project: Training Seminar - 3-9 November 1990

Nov/Dec. 1990 First administration of tests and observations

May/Oct. 1991 Second administration of the tests and observations

January 1992 Third Consultancy Visit by CJW 4-11 January 1992

May 1992 Third administration of the tests and observations

June 1992 Fourth Consultancy Visit by CJW 15th-19th June 1992

July-Sep 1992 Attachment of 3 Guinean team members to CALS to write up the study

ANNEX TABLES: RESULTS

Table A1: Numbers of Students

A1a: Control Group

Visit	Dictation	Gap-filling	Writing
1	288	281	289

2	204	204	201
3	98	115	116
2-1	193	194	185
3-1	96	113	113
2-3	72	86	86

A1b: Experimental Group

Visit	Dictation	Gap-filling	Writing
1	432	413	425
2	286	280	290
3	142	166	172
2-1	280	267	277
3-1	141	161	164
2-3	130	152	155

The numbers in the comparison between visits 3 and 2 and between visits 3 and 1 are different because some students were present at the second visit but not at the first and vice versa. Thus the two sets of improvement in scores are not directly comparable and must be considered separately.

Table A2: Gap-Filling Test Results

A2a: Control Teachers

	Variable	Mean	Std Dev	Prob> T
visit 1	(i1-i34)	7.84	2.15	0.0001
visit 1	(i35-i70)	4.01	2.56	0.0015
visit 1	(i1-i70)	11.85	4.67	0.0001
visit 2	(i1-i34)	12.83	2.51	0.0001
visit 2	(i35-i70)	10.04	4.26	0.0001
visit 2	(i1-i70)	22.87	6.69	0.0001
visit 3	(i1-i36)	16.65	2.49	0.0001
visit 3	(i37-75)	6.96	3.58	0.0051
visit 3	(i1-75)	23.62	5.68	0.0002

vis2-vis1	(i35-i70)	5.70	4.03	0.0028
vis2-vis1	(i1-i70)	10.45	6.04	0.0008
vis3-vis1	(i35-i70)	12.09	3.18	0.0002
vis3-vis2	(i35-i70)	5.22	4.80	0.0446

A2b: Experimental Teachers

	Variable	Mean	Std Dev	Prob> T
visit 1	(i1-i34)	8.39	2.01	0.0001
visit 1	(i35-i70)	3.45	1.40	0.0001
visit 1	(i1-i70)	11.85	3.20	0.0001
visit 2	(i1-i34)	12.73	2.30	0.0001
visit 2	(i35-i70)	9.42	2.86	0.0001
visit 2	(i1-i70)	22.16	4.63	0.0001
visit 3	(i1-i36)	16.52	5.02	0.0001
visit 3	(i37-75)	6.88	4.77	0.0004
visit 3	(i1-75)	23.41	9.67	0.0001
vis2-vis1	(i35-i70)	5.77	3.26	0.0001
vis2-vis1	(i1-i70)	9.88	5.56	0.0001
vis3-vis1	(i35-i70)	12.70	5.68	0.0001
vis3-vis2	(i35-i70)	6.92	3.60	0.0001

Table A3 Dictation Test Results

A3a: Control Teachers

	Variable	Mean	Std Dev	Prob> T
visit 1		30.89 (100)	8.79	0.0001
visit 2		43.91 (100)	10.38	0.0001
visit 3	(i1-i43)	44.17 (100)	5.34	0.0001
visit 3	(i44-i53)	10.98 (100)	6.28	0.0078
visit 3(all items)		55.16 (200)	11.20	0.0001
visit2 - visit1		12.63 (100)	5.51	0.0001
visit3 - visit1		14.39 (100)	12.44	0.0366
visit3 - visit2		-1.51 (100)	11.57	0.7609

A3b: Experimental Teachers

	Variable	Mean	Std Dev	Prob> T
visit 1		29.10 (100)	6.42	0.0001
visit 2		44.75 (100)	6.99	0.0001

visit 3	(i1-i43)	46.95 (100)	9.75	0.0001
visit 3	(i44-i53)	11.00 (100)	5.93	0.0001
visit 3(all items)		57.95 (200)	14.57	0.0001
visit2 - visit1		14.09 (100)	7.83	0.0001
visit3 - visit1		15.32 (100)	9.22	0.0001
visit3 - visit2		0.95 (100)	9.37	0.7300

Table A4 Writing Test Results

A4a: Control Teachers

	Variable	Mean	Std Dev	Prob> T
visit 1	(i1-120)	17.97 (100)	8.89	0.0003
visit 2	(i1-i20)	32.40 (100)	12.08	0.0001
visit 3	(i1-i20)	38.47 (100)	9.73	0.0002
visit 3	(i21-i35)	15.84 (75)	4.89	0.0005
visit 3(all items)		54.31 (175)	13.52	0.0002
visit2 - visit1		13.95 (100)	13.14	0.0129
visit3 - visit1		20.81 (100)	9.00	0.0024
visit3 - visit2		-0.16 (100)	14.14	0.9780

A4b: Experimental Teachers

	Variable	Mean	Std Dev	Prob> T
visit 1	(i1-120)	17.48 (100)	6.70	0.0001
visit 2	(i1-i20)	26.16 (100)	6.44	0.0001
visit 3	(i1-i20)	35.79 (100)	10.43	0.0001
visit 3	(i21-i35)	19.86 (75)	9.59	0.0001
visit 3(all items)		55.65 (175)	19.46	0.0001
visit2 - visit1		8.10 (100)	8.34	0.0044
visit3 - visit1		17.03 (100)	10.71	0.0002
visit3 - visit2		8.63 (100)	7.48	0.0021

Table A5 Oral Test Results

A5a: Control group - Raw Scores

Test	Visit	N	Mean	Std Dev	T	Prob> T
Imitation	1	9	5.907(17)	1.085	16.32	0.0001

	2	9	6.338(17)	1.552	12.24	0.0001
	3	6	7.375(17)	2.441	7.39	0.0007
	2-1	9	0.122	1.155	0.31	0.7591
	3-1	6	1.347	2.698	1.22	0.2759
	3-2	6	1.055	2.642	0.97	0.3729
Interview	1	9	3.168(10)	1.096	8.66	0.0001
	2	9	5.459(10)	1.541	10.62	0.0001
	3	6	5.916(10)	2.021	7.16	0.0008
	2-1	9	2.009	2.191	2.75	0.0250
	3-1	6	2.861	2.101	3.33	0.0207
	3-2	6	0.527	1.495	0.86	0.4270
Inform.	1	9	2.806(10)	1.445	5.82	0.0004
Transfer	2	9	3.446(10)	1.607	6.43	0.0002
	3	6	6.333(10)	2.812	5.51	0.0027
	2-1	9	0.625	1.892	0.99	0.3507
	3-1	6	3.944	2.245	4.30	0.0077
	3-2	6	2.694	3.190	2.06	0.0934

A5b: Experimental Group - Raw Scores

Test	Visit	N	Mean	Std Dev	T	Prob> T
Imitation	1	13	6.402(17)	1.348	17.12	0.0001
	2	13	6.406(17)	1.670	13.82	0.0001
	3	11	8.500(17)	2.083	13.53	0.0001
	2-1	13	0.208	1.679	0.44	0.6616
	3-1	11	1.530	1.749	2.90	0.0158
	3-2	10	1.366	2.584	1.67	0.1288
Interview	1	13	3.976(10)	1.000	14.33	0.0001
	2	13	6.089(10)	0.792	27.68	0.0001
	3	11	7.727(10)	1.325	19.33	0.0001
	2-1	13	2.164	0.736	10.58	0.0001
	3-1	11	3.621	1.651	7.27	0.0001
	3-2	10	1.433	2.130	2.12	0.0623
Inform.	1	13	2.848(10)	1.718	5.97	0.0001
Transfer	2	13	4.202(10)	1.729	8.76	0.0001
	3	10	6.866(10)	3.551	6.11	0.0002
	2-1	13	1.374	1.424	3.47	0.0046
	3-1	10	3.450	2.627	4.15	0.0025
	3-2	8	2.270	3.981	1.61	0.1507

Results from tracer study questionnaire

Table A6a: Percentages of Respondents Concerning the Importance of Training Programme Elements.

	Very Important	Important	Very Important/Important	Not Important	Don't Know
Language development	57.1	26.7	83.8	0.0	16.2
Literature	37.5	25.0	62.5	0.0	37.5
Methodology	46.9	37.5	84.4	0.0	15.6
In-House Prac. Work	25.0	50.0	75.0	9.4	15.6
Teaching Prac. (Lycees)	56.2	22.0	78.2	6.2	15.6
Written Assignments	75.0	12.5	87.5	0.0	12.5
Other Components	50.0	29.1	79.1	8.3	12.5

This summarises the returns to the individual elements of each section of the course.

Table A6b: Percentages of Respondents Concerning the Sufficiency of Training Programme Elements.

	More than Sufficient	Sufficient	More than Sufficient/Sufficient	Less than Sufficient	Don't Know
Language development	14.3	26.7	41.0	32.1	26.7
Literature	0.0	37.5	37.5	25.0	37.5
Methodology	1.5	39.0	40.5	32.8	26.5
In-House Prac. Work	0.0	40.6	40.6	31.2	28.1

Teaching Prac. (Lycees)	12.5	34.4	46.9	25.0	28.0
Written Assignments	12.5	25.0	37.5	37.5	25.0
Other Components	4.1	25.0	29.1	45.8	25.0

This summarises the returns to the individual elements of each section of the course.

Table A7a : Number of Responses Concerning the Balance within the Training Programme and the ENS Methods of Teaching.

	Less than good	Good	More than good	Don't know
Question 1	1	2	1	4
Question 3	0	6	2	0

N.B.-Q1 : What do you think of the balance between language development and professional training ?

Q3 : Do you have any comments on the ENS method of teaching?

Table A7b: Number of Responses Concerning the Usefulness of Resources and the Bo Seminar.

Q2: Do you have any comments about the resources made available to you?

a) books, including textbooks,simplified readers, general fiction, journals.

b) cassette players

c) video

Q4: Please, give your assessment of the value of the Bo seminar.

	Less than useful	Useful	More than useful	Don't know
Question 2a)	0	4	4	3
b)	1	2	1	4
c)	2	2	1	3
Question 4	0	1	7	1

Question 5 Special Training Needs Expressed by Respondents

Q5: Were there any gaps in the ENS programme which you are now able to

identify ? -Identify your five biggest problems in teaching which the Maneah training programme did not sufficiently prepare you for.

Use of French :	1
Large classroom management:	2
Lack of teaching materials:	1
Teaching Oral:	1
Assessing Oral:	1
Teaching Reading:	1
Teaching Listening:	1
Teaching Pronunciation:	1
Teaching grammar:	1
Evaluation:	1

(3 don't know)

Q6: Please comment on any other points not raised above, or covered inadequately.

Require Further training: 4 (4 don't know)