DOCUMENT RESUME

ED 045 969 AL 002 694

AUTHOR	Lindblad, Torsten; Levin, Lennart
TITLE	Teaching Grammar. An Experiment in Applied Psycholinguistics. Assessing Three Different Methods of Teaching Grammatical Structures in English as a Foreign Language.
INSTITUTION	Gothenburg School of Education (Sweden).; Gothenburg Univ. (Sweden). Dept. of English.
SPONS AGENCY	Swedish National Board of Education, Stockholm.
REPORT NO	GUME-P-4
PUB DATE	Dec 70
NOTE	186p.; Appeared as "Research Bulletin No. 6, December 1970" of the Gothenburg School of Education
AVAILABLE FROM	Department of Educational Research, Gothenburg School of Education, Ovre Husargatan 34, S-413 14 Gothenburg, Sweden
EDRS PRICE	EDRS Price MF-$0.65 HC Not Available from EDRS.
DESCRIPTORS	Contrastive Linguistics, *Educational Experiments, *English (Second Language), *Language Instruction, Language Tests, Pattern Drills (Language), *Psycholinguistics, Research Projects, Statistical Studies, Student Attitudes, *Teaching Methods

ABSTRACT

Earlier reports from the GUME project (Goteborg, UndervisningsMetod i Engelska-Gothenburg/ Teaching/Methods/English) have shown that different methods of teaching English produce no significant differences in learning effects. In the present study, a direct continuation of earlier ones, modifications in design, teaching strategies, etc., were made in order to increase the probability of detecting true differences between methods, if such existed. Three methods were compared: the Implicit method, the Explicit-English method(Ee), and the Explicit-Swedish method(Es). All three have systematized drills, but Ee and Es feature analysis and explanations as well; in Ee such explanations are given in English, in Es in Swedish. The present study differed from earlier ones in several ways: a new type of explanation was used, the duration of the experiment was prolonged, and the grammatical content was more varied. Main effects were investigated by analysis of covariance and interaction effects by analysis of variance. Individual scores were used as units of analysis, and various measures of progress during the experiment were used in the comparisons. Results agreed with earlier findings that the three different methods do not generate any differences in learning effects. See related document FD 038 172. [Not available in hard copy because of marginal legibility of the original document.] (Author/PWB)

T E A C H I N G G R A M M A R

An Experiment in Applied Psycholinguistics

Assessing Three Different Methods of Teaching Grammatical Structures in English as a Foreign Language

by

Torsten Lindblad and Lennart Levin

Gothenburg School of Education
Department of Educational Research

and

Department of English

University of Gothenburg
Department of English

December, 1970

- He's changed to the latest model.
- Car?
- No, method.

Within the GUME project (Göteborg, UndervisningsMetod i Engelska = Gothenburg/Teaching/Methods/English) earlier studies showed no significant differences in learning effects between different methods of teaching English.

The present study is a direct continuation of the earlier studies. Modifications in design, teaching strategies, etc., were made in order to increase the probability of detecting true differences between methods, if such existed. As in the previous experiments, the three methods being compared were: the Implicit method, the Explicit-English method, and the Explicit-Swedish method. In all the methods the students have systematized drills; in Ee and Es the students have analysis and explanations as well. In Ee these explanations are given in the target language and in Es in the source language. In Es comparisons are also made with the corresponding grammatical structures in Swedish.

In comparison with earlier investigations, the present study - GUME 4 - was modified in the following respects: a new type of explanation was used, the duration of the experiment was prolonged, the grammatical content was more varied, the study was carried out at another grade level, and the teachers did take a limited part in the teaching procedure.

Main effects were investigated by analysis of covariance and interaction effects by analysis of variance (two-way classification). Individual scores and, in one case, school class means were used as units of analysis. Various measures of progress during the experiment were used in the comparisons.

ACKNOWLEDGEMENTS

The GUME Project is an interdepartmental undertaking, carried out in cooperation between the English Departments of the University of Gothenburg and of the Gothenburg School of Education (Lärarhögskolan i Göteborg) and the Department of Educational Research of the School of Education. Professors Alvar Ellegård, the University English Dep't, and Karl-Gustaf Stukåt, Educational Research Dep't, have acted as research consultant. The project has been sponsored financially by the National Board of Education, Bureau L 4.

The writers of this report want to express their gratitude to those who have helped in various ways. First and foremost we want to express our indebtedness to Ingvar Carlsson who carried a heavy load in the execution of the project but who has not, because of other duties, been able to assist us in the last phase of the work.

We want to thank all those teachers in Gothenburg who made it possible for us to carry out the experiments by their willingness to participate even though this meant extra work. We also appreciate the cooperation of the Gothenburg Board of Education for understanding and help with technical problems.

Two other people should be mentioned: Per Högberg, our computer and programming expert, whose help has been very valuable, and Kerstin Davidsson, our patient and skilful typist who has done an excellent job in spite of a difficult manuscript.

We have been in a position to discuss various phases of the project with many friends and experts in different fields. To all of them our sincere gratitude!

Gothenburg, December, 1970.

Torsten Lindbiad Lennart Levin

TABLE OF CONTENTS

ACKNOWLEDGEMENTS i

TABLE OF CONTENTS ii

List of Figures vi

List of Tables vii

INTRODUCTORY NOTE ON THE TREATMENT OF STATISTICS ix

BACKGROUND 1

Earlier GUME Activities 1

Modifications of Earlier Designs 3

Total GUME Activities 6

THE PENNSYLVANIA PROJECT CONTINUED 8

EXPERIMENTATION IN A FIELD SETTING - SOME REFLEXIONS 18

Comparative Experiments - Pros and Cons 18

The GUME Project - Some Comments 20

Hypothetical Treatment Effects 22

METHODS IN FOREIGN LANGUAGE TEACHING 24

Introduction 24

Some Well-known Methods 24

The Authorized Curriculum for Swedish Schools 25

GUME Methods 26

Implicit 27

The Explicit Methods 28

Current Debate 30

A DESCRIPTION OF THE LESSON SERIES 32

Schedule 32

Contents 32

One Lesson Described 34

Explicit Lessons 37

"Ee 7" 40

THE GUME 4 PROJECT - A DESCRIPTION OF THE LAY-OUT 41

Objectives 41

Experimental Procedure 41

Schedule 41

Technical Arrangements 42
Teaching Methods 42
The Experimental Population 42
Number of School Classes 42
Selection of School Classes 43
Assignment to Treatments 44
Drop-out Rate 44

EVALUATION INSTRUMENTS 45
The Achievement Test 45
The Twelve Parts of the Test 46
Marking and Test Administration 47
Validity and Reliability 48
The Pupil Attitude Test 49
The Teacher Attitude Test 50
The Standardized Test 50
PACT 52
The Intelligence Test 52
Other Measures 54
Social Class 54
Grades 54

DATA TREATMENT 55
The Statistical Program 55
Experimental Design 55
Computation of Main Effects 56
Computation of Interaction Effects 57
Retention 57
Various Measures of Progress 57

STATISTICAL DESCRIPTION OF THE EXPERIMENTAL POPULATION 59
Attendance 59
Assignment to Treatments 59
Social Class 60
Course Choice 61
Representativity of the Experimental Group in Certain Variables 62
IQ 62
Grades 64
The Standardized Test in English 65
Characteristics of the Treatments Groups 66

MAIN RESULTS 67
Overall Progress during the Experiment 67
Progress - Main Effects 69
School Class Progress 69
Individual Progress per Method 69
Progress - Interaction 72
Progress - Main Effects at Two Ability Levels 74
Post-test - Main Effects 75
Post-test - Interaction 75
Progress - the School Class Mean as the Unit of Analysis . 77
Additional Studies of Progress 77
Drop-outs 80
Some Findings in Social Group "0" 81
Findings Related to Course Choice 82
Results on the Different Part Tests 87
Method Differences 89
"Critical Items" 89

CORRELATION STUDIES 93
Background Variables 93
Social Class 93
Pupil Attitude 93
Intelligence 93
Grades 95
The Standardized test 95
PACT 95
Correlations with Course Choice 96
Pre-test and Post-test Correlations 97
Progress Correlations 100

ATTITUDES 103
Pupils' Attitudes to the Project (General) 103
Pupils' Attitudes to the Explanations 105
Interest in English 108
Teachers' Attitudes 109
General methods questions 109
Questions on the project 110

DISCUSSION OF RESULTS 113

THE GUME 1 AND 2 FOLLOW-UP STUDIES 1969/70 118
Frequency Distribution 123
Some Correlations 127

SUMMARY 130

BIBLIOGRAPHY 134

APPENDICES A - F

List of Figures.

Fig. 1: A Survey of GUME Research Activities 1968-1971 7

Fig. 2: "The Methods Continuum"..... 25

Fig. 3: Pennsylvania and GUME Methods Compared 30

Fig. 4: Two Pages from the Pupils' Booklet 36

Fig. 5: Example of Page in PACT 53

Fig. 6: Distribution of Individual Progress 68

Fig. 7: The Progress on the 27 Experimental Classes 70

Fig. 8: Distribution of sk/ak Choices per Social Class 83

Fig. 9: Distribution of Scores on the IQ Test 85

Fig. 10: Distribution of Scores on PACT 85

Fig. 11: Distribution of Scores on the Standardized English Test 85

Fig. 12: Progress in One School Year 124

Fig. 13.a & b:The GUME I Follow-up: Distribution of Scores on a) the Pre-test, b) the Post-test 126

Fig. 14 a & b: The GUME II Follow-up: Distribution of Scores on a) the Pre-test b) the Post-test 126

List of Tables

Table 1:	Total Running Time of the Twelve Lessons in the Three Methods	33
Table 2:	Grammatical Structures and When They Were Practised	34
Table 3:	Exact Times for Explanations in the E Groups	38
Table 4:	Reliability Coefficients for the Achievement Test	49
Table 5:	Analyses of Covariance Performed	56
Table 6:	Attendance of the Experimental Population during the Series of Lessons	59
Table 7:	Distribution of Pupils on Teaching Methods	60
Table 8:	Distribution according to Social Class	60
Table 9:	Distribution according to Social Class, Percentages	61
Table 10:	Distribution according to Course Choice in English for Grade 7	62
Table 11:	Means and Standard Deviations on the DBA-test	63
Table 12:	Grades: means and standard deviations	64
Table 13:	The Standardized Test: means and standard deviations	65
Table 14:	Analyses of Variance (one-way) of Certain Background Variables	66
Table 15:	Pre-test, Post-test and Progress: Means and Standard Deviations	67
Table 16:	Analyses of Covariance (PROGRESS)	72
Table 17:	Analysis of Variance (two-way) (PROGRESS)	73
Table 18:	Analyses of Covariance at Two Levels of Ability (Upper and Lower) (PROGRESS)	74
Table 19:	Analysis of Covariance (POST-TEST)	75
Table 20:	Analysis of Variance (two-way) (POST-TEST)	76
Table 21:	Analysis of Covariance of School Class Means (POST-TEST)	77
Table 22:	School Class Means on the Variable (actual/possible improvement x 100)	78

Table 23: Analysis of Variance (one-way) of School Class Means on the Variable (actual/possible improvement x 100) 79

Table 24: School Class Means on the Variable (Post-test - Pre-test/Pre-test) x 100 79

Table 25: Means and Standard Deviations for the Experimental Population and Drop-outs 80

Table 26: Means and Standard Deviations for the Experimental Population and Social Group "0" 81

Table 27: Distribution of Social Class in Relation to Course Choice 82

Table 28: Means and Standard Deviations for Presumptive sk and ak Pupils 83

Table 29: Results on the Parts of the Tests per Method 88

Table 30: So-Called Critical Items per Method 91

Table 31: Intercorrelations between Main Variables 94

Table 32: Correlations between the Pupils' Course Choice for Grade 7 and Certain Other Variables 97

Table 33: Pre- and Post-test Correlations with Certain Other Variables 98

Table 34: Pre- and Post-test Intercorrelations 99

Table 35: Progress Correlations 101

Table 36a: Pupils' Attitudes (Explanations excluded) 104

Table 36b: Pupils' Attitudes to Explanations 107

Table 37: Interest in English 108

Table 38: GUME 1: Follow-up and Original Results 120

Table 39: GUME 2: Follow-up and Original Results 120

Table 40: GUME 1: Follow-up and Original Results for ak and sk 122

Table 41: GUME 2: Follow-up and Original Results for ak and sk 122

Table 42: GUME 1: Test Correlations 128

Table 43: GUME 2: Test Correlations 128

Table 44: GUME 1 and 2 Follow-up Pupils: Progress Correlations 129

INTRODUCTORY NOTE ON THE TREATMENT OF STATISTICS

The study dealt with in the present report is an interdepartmental (tvärvetenskaplig) undertaking, one of the authors representing English as an academic discipline and a school subject, one representing pedagogy as an academic discipline and educational research and statistics as theoretical background. We have written the report with two quite distinct groups of readers in mind: teachers of English and educational researchers. The former group normally has little training in statistics and has a tendency to shy away from figures, the latter has training in this field and is perhaps more used to reading reports like the present one. This has caused problems in writing the report.

What we have tried to do is the following. We have used ordinary statistical methods and give as much information and as many tables as will hopefully satisfy the second group of our intended readers. But we have also tried to arrange the tables so as to facilitate the reading of them for the first group of readers. The language teacher with little training in statistics is recommended to study columns and tables of means and standard deviations, and N (see below). In commenting on our tables we have not always limited ourselves to conclusions and discussions of these but have also tried to explain how we arrived at these conclusions, how the figures ought to be understood, what size a certain figure must reach to be "significant", etc. We hope that those readers who find these comments superfluous will understand the pedagogical *raison d'être* for them and will just skip them.

For the convenience of the reader with little statistical training some frequent symbols and terms are explained below. In almost every case the explanation is an attempt at giving the general idea or practical use of a symbol rather than an adequate or in all respects logical definition of it.

N — The *number* of pupils for which a certain measure is given.

$\bar{x}$ — The *arithmetic mean* of a group.

s — The *standard deviation*, i.e. a measure of the extent to which the scores for a certain group vary. The larger the s, the *more heterogeneous the group*. A single s does not carry much meaning; the measure should be used for comparison with other s's.

t	This value indicates whether a difference between the means of two groups is "statistically significant" or whether it can be explained as a chance occurrence. As f as the analyses in the present report are concerned the critical t-value is 1.96, i.e. when t is equal to or gre ter than 1.96, the difference under investigation is cor sidered a real, non-chance difference.
F	F, or the F-ratio, is used for the same purposes as t. However, F is the relevant characteristic when more thar two means are compared. Since t h r e e teaching methods are being compared in the present study, F appea quite often in our tables. The corresponding critical value for interpreting differences as true differences around 3.00; this figure varies a little depending on th number of pupils.
T-scale	A scale with a theoretical mean of 50 and a standard dev tion of 10. The scores on a certain test, whatever its and s, can be transformed into T-scores.
Stanine scale	A 9-point scale with a theoretical mean of 5 and a stan ard deviation of 2. In contrast to the T-scale, the stanine scale has a so-called standardized (normalized) distribution of scores. Scores on a test may be transfo to stanines by giving the top and bottom 4 % of the pup 9 and 1 points respectively, the next 7 % at each end 8 and 2 respectively, thus: 9 (4 %), 8 (7 %), 7 (12 %), 6 (17 %), 5 (20 %), 4 (17 %), 3 (12 %), 2 (7 %), 1 (4 %
Analysis of variance	The method is used for comparing the means of three or m groups which have been exposed to different treatments. the groups respond in different ways, i.e. are their mea statistically different? In this sort of analysis, the variation in scores b e t w e e n groups and w i t i n groups are considered in relation to each other. F true differences between group means to exist, it is necessary for the variation in scores *between* groups to be greater than the variation *within* groups. This sort o analysis yields an F-ratio (see above).
Analysis of covariance	The same as the above method with the addition that the groups' standing on essential background variables is taken into account. For instance, if three groups are to be compared with respect to learning effects and the groups differ substantially in intelligence, it is very probable that the group having the brightest children (a not necessarily the children exposed to the "best" metho would come out as the best. In an analysis of covariance differences of this sort are equalled out statistically. This analysis also yields an F-ratio.

Adjusted means Refers to analyses of covariance. The means of the groups being compared are adjusted for variation between the groups in background variables. Briefly, if three groups were to rank A > B > C in a teaching experiment and their values in the background variable, say intelligence, also ranked A > B > C, the adjusted means would be equal for the three groups. Thus, when original differences between the three groups were taken into consideration, differences obtained after the teaching experiment disappeared.

X^2 (Chi^2) A value used to indicate whether the answers on, for instance, a questionnaire are evenly distributed among the response alternatives. It is used to investigate if the particular distribution of answers (given by a group of individuals) is in accordance with an expected distribution and if a deviation in this respect is so small that it might be explained as a chance occurrence. The differences between observed and (theoretically) expected frequencies add up to a so-called X^2 -value; the higher this value, the more probable is the conclusion that the group (of pupils, etc) under consideration deviates significantly from "the norm".

BACKGROUND

Earlier GUME Activities

The present report describes further research on the teaching of English as a foreign language by members of the so-called GUME project. The work should be viewed against the background of four separate reports, published in 1969 (see special section of the bibliography, page 134) and describing teaching method comparisons performed thus far. For readers not familiar with the publications just mentioned, a brief resumé may be in order:

Three parallel studies, identical in design, were carried out in order to investigate three different methods of teaching grammatical structures in English as a foreign language. The studies were performed during the autumn term of 1968 and the spring term of 1969. Three different areas of English grammar that are known to cause Swedish students difficulty were selected for investigation:

GUME 1 The do-construction
GUME 2 The some-any dichotomy
GUME 3 The passive voice

The three methods of instruction (independent variables) investigated in each of the experiments were:

Im *The Implicit method*, where the students had systematized drills but no analysis or explanations of the grammatical structures involved.

Ee *The Explicit-English method*, where the students had systematized drills and, in addition, analysis and explanations in the target language (English). The time allotted to the explanations was taken from the drills.

Es *The Explicit-Swedish method*, where the students had systematized drills and, in addition, analysis and explanations in the source language (Swedish); comparisons with corresponding structures in Swedish were also made. The time allotted to the explanations was taken from the drills.

In each part project 18 school classes took part, 6 per teaching strategy. Of these 6 classes, 4 represented the advanced course (särskild kurs, abbreviated sk) and 2 the easier course (allmän kurs, abbreviated

ak). Thus the total GUME project contained 54 classes, of which 36 were in sk and 18 in ak. The school classes, representing a wide geographical variation within the Gothenburg area, were randomly assigned to the teaching methods.

For each part project 3 lesson series (Im/Ee/Es) were constructed, each consisting of 6 lessons. In order to control the teacher factor "canned" lessons were used throughout the experiment. The students listened to the programs via headsets with induction receivers. Magnetic wires were installed and tape-recorders used in every classroom; this simple arrangement comes close to a language lab as far as sound quality is concerned.

Within each part project, the pupils' progress was measured by an achievement test, designed to correspond to the specific objectives of the part project in question. That is to say, the same test was administered as Pre-test before and as Post-test after the experiment, the difference between the two being the Progress score for each pupil. The identical test was also administered as Re-test approximately one month after the experiment in order to measure retention.

The pupils' attitudes to various aspects of the study were collected by means of a questionnaire.

Since the treatment groups within each experiment were not experimentally controlled, statistical control was undertaken by means of analysis of covariance. The covariates resorted to were "general intelligence" (the verbal, inductive and spatial factors of an IQ test frequently used in Swedish schools), grades in English, Swedish and Mathematics, and in some analyses Pre-test scores. Partly the analyses were made with Progress scores as the dependent variable and partly with Post-test scores as the dependent variable.

In the various statistical analyses the experimental population was divided according to two principles: in one type of analysis sk and ak were treated separately, in another the population was divided into three equal parts according to IQ scores, the Upper, Middle and Lower third. In the latter case analyses of variance (two-way classification) were performed in order to investigate interaction between ability level and teaching method.

More detailed information about the statistical treatment of GUME 1-3 will not be given in this connection, suffice it to say that a total of 60 (sixty) analyses of covariance and variance were performed.

In *two* of them statistically significant differences were obtained, which is less than could be accounted for by mere chance even if the null hypothesis (no difference between treatments) were true. Nor was there any evidence of interaction between ability level and teaching strategy in the study.

Thus the GUME 1-3 experiments have not shown that any differences are produced by the three teaching methods.

It is sometimes argued that "insignificant" results like those obtained in GUME 1-3 have low social utility (Anderson, 1969) since they do not provide much support for people involved in production of teaching materials.

In the three studies referred to, however, the main concern was with the basic problem of whether explanations facilitate learning rather than with production of materials. Consequently the lessons were designed to provide an answer to the basic research question without necessarily coming close to "ordinary" lessons. Even so, no differences were found between the three teaching methods compared. (If significant differences had appeared, they would still have been of limited interest *with respect to the production of materials.*)

Findings like those just reported are not uncommon in educational research (Stephens, 1967). True differences between methods may have escaped detection because the experiments lacked statistical power (Stanley, 1970) or because of deficiencies in the planning and execution of the studies. There is also the possibility that no true differences between the methods exist, though this can never be proved.

Modifications of Earlier Designs.

When the present experiment was planned, the teaching strategies and general design were modified in essential respects to increase the probability of detecting differences, if such existed. The teaching strategies, the lessons, and the experimental procedure will be described in detail later; here we shall only give a brief description of the modifications alluded to above.

1. In GUME 1-3 great effort was made (in Ee and Es) to keep the time allotted to explanations in each lesson constant. Furthermore it was judged essential that the explanations be of substantial length; in fact, the explanation time approximated 1/3 of the lesson time in Ee and Es. However, pupils' questionnaires as well as observation of

classroom activities suggested that the explanations were too long. As a result, some experimental strictness (= equality of explanation time in Ee and Es) was sacrifized for the benefit of "optimal" explanations. This perhaps somewhat pretentious term indicates that the explanations were introduced "when they were needed" and in a way judged relevant with regard to optimal learning. As it appeared, this strategy had the effect that the explanations usually became shorter and that the Ee and Es explanations could, and did, vary in length.

2. A common feature of comparative field studies is their relatively short duration. It is the exception rather than the rule that the treatments are applied for any considerable amount of time, for instance a school term or more. Although this would be desirable in most cases, practical and monetary considerations usually restrict the researcher's actions. As was mentioned earlier, GUME 1-3 consisted of 6 lessons, which was what the resources permitted at that time. The present study, GUME 4, consisted of 12 lessons, administered during one month. Although this may still be considered a relatively small amount of time for a treatment to show its potential, there should be reasonable probability for true differences to appear. Besides, even an experiment consisting of as few as 12 "canned" lessons, or rather 12 per method, i.e. 3 x 12 lessons, as was the case in GUME 4, takes a considerable time to prepare and administer.

3. In GUME 1-3 the three part projects concentrated on one syntactic structure each. In GUME 4 it was thought desirable to expose the students to a somewhat wider range of grammatical structures or problems, thereby creating greater variety and, hopefully, higher motivation, and also increasing the probability of detecting method differences. The particular grammatical items chosen will be presented in due course.

4. The GUME 1-3 experiments were performed in grade 7, i.e. the first grade of the Upper stage of the Swedish comprehensive school, where the pupils take two separate courses in English. The present study was carried out in grade 6. One reason for moving to grade 6 is the fact that there the pupils take a number of standardized achievement tests in English, which might be used for the purpose of treatment group comparisons and description of the experimental population. Another not unimportant advantage of performing the study

in grade 6 is the class-teacher system prevalent there, which means that practical problems (disturbances in research schedule because of unforeseen circumstances, etc.) can be more easily solved than in classes at the Upper stage where a number of teachers will be affected by such changes.

5. In GUME 1-3 assistants administered the lessons, i.e. their sole function was to start the tape and hand out the booklets containing the lesson material. Observation of classroom activities revealed, however, that in some classes the pupils did not take a very active part in the oral drills. The assistants were instructed not to interfere in the teaching procedure; thus nothing prevented the pupils from being inactive. Although the idea behind using "canned" lessons is to control the teacher factor, it was judged preferable in GUME 4 to let the live teacher control pupil activities with respect to oral drills. Thus the teachers were instructed to activate the pupils' repeating after the tape and to indicate, by pointing, etc., which of the pupils should answer a question. This participation by the teachers was thus intended as a check on pupil activities and should, if carried out according to instructions, be almost identical among the teachers. However, variation in teacher behaviour should be taken into account as a possible source of error in the experiment.

The above modifications, compared with earlier research within the GUME project, are all aimed at increasing the internal as well as the external validity of the experiment. Thus in GUME 4 (as opposed to GUME 1-3):

1. "Optimal" explanations are used
2. The duration of the experiment is doubled
3. More grammatical structures are taught
4. The study is carried out in grade 6
5. The ordinary teacher administers the "canned" lessons.

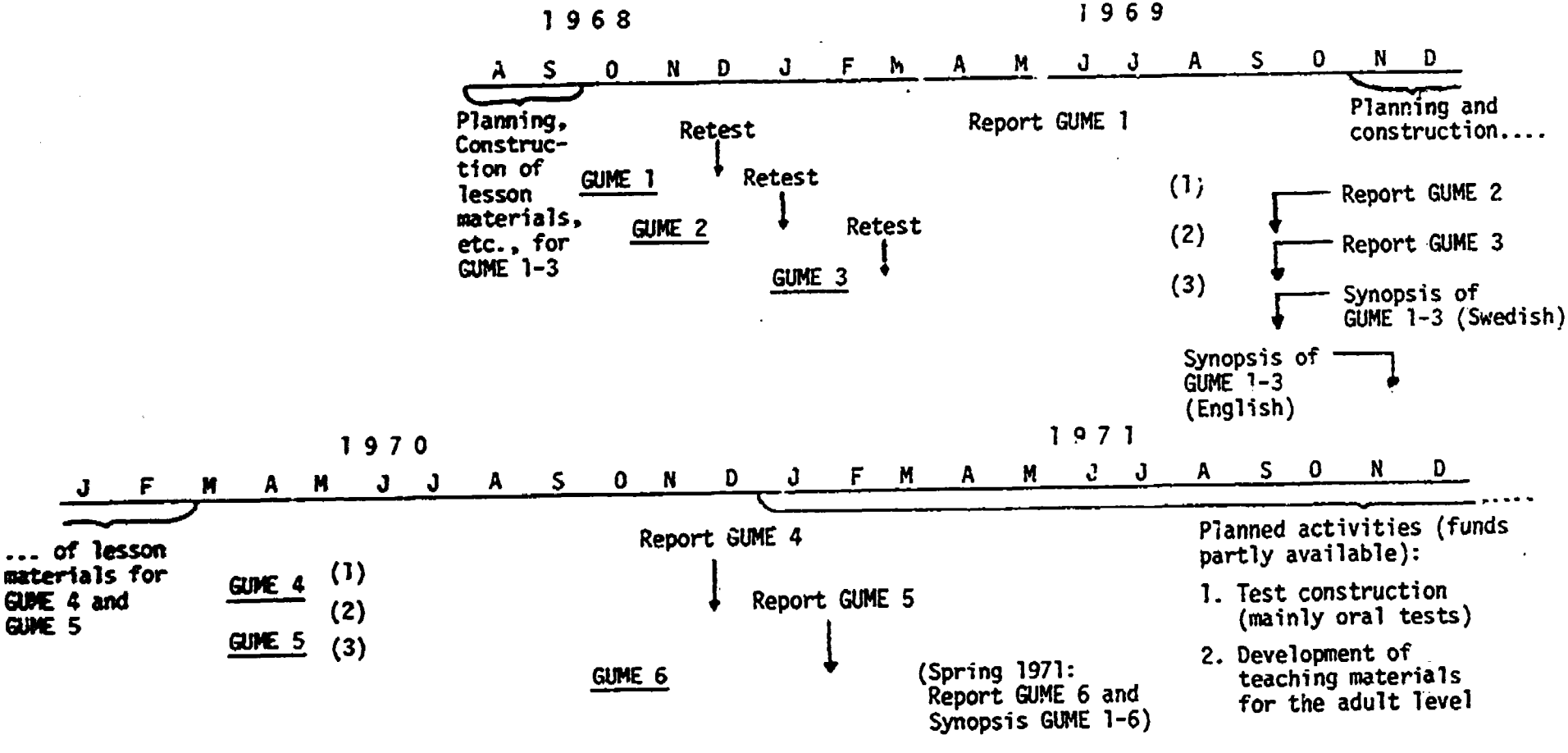
Total GUME Activities.

So far the reader has become acquainted with the three first part projects, GUME 1-3. As has been shown, the results generated some hypotheses about new directions for further research to take. In the case of the present study, GUME 4, the revised research strategy has been presented in the preceding section. However, two more part projects, GUME 5 and GUME 6, were started during 1970 in order to provide further knowledge within the field of foreign language teaching. The two part projects will be presented in forthcoming reports (GUME 5 in January, 1971, and GUME 6 in April/May, 1971). The following brief discussion of the two studies is intended to complete the picture of the total GUME activities.

GUME 5 was carried out simultaneously with the present study though in grade 8. It is a direct continuation of GUME 3 as far as lesson content is concerned. The passive voice is the syntactic structure taught and the same pedagogical expert is responsible for the production of teaching materials. The pupils in grade 8 take two separate courses in English. One and the same teaching program was used in both courses. Finding out how this functioned has become even more interesting after the introduction of the new Curriculum for Swedish Schools (Lgr 69) which states that the same objectives should apply to both courses.

GUME 6 is undertaken at the adult level. The strategy adopted in this case is to compare two methods only, one of an audiolingual kind with numerous structure drills and no explanations, and one with very few drills but with explanations in the source language. The two methods are intentionally made more distinct than for instance Im vs. Ee/Es in the earlier GUME experiments. Fig. 1 gives a survey of the GUME studies, performed as well as planned. At one point a clarification is necessary; the figures 1, 2 and 3, appearing in two positions, indicate that the achievement tests used in GUME 1, 2 and 3 respectively were administered in control classes at the beginning and the end of the school year. The purpose was to find out to what extent the structures taught during the GUME experiments are actually learnt in a school year without the teachers' paying special attention to those structures. Progress in the control classes will be commented on in the present report (p. 118 ff).

Figure 1: A Survey of GUME Research Activities 1968-1971.



THE PENNSYLVANIA PROJECT CONTINUED

The largest undertaking in recent years in the field of educational research concerning the teaching of foreign languages is the Pennsylvania study. The GUME project is a similar enterprise although on a much more modest scale, smaller in scope and personnel. We have studied the Pennsylvania reports carefully and tried to learn both from those parts of the design and evaluation which are worthy of imitation, and from the mistakes and shortcomings. In an earlier report (Levin, 1969, p. 6 ff) we gave a commented outline of the study, including what had been reported by September, 1969. The debate in USA has been lively, and since much of the criticism levelled at the Pennsylvania Project might be directed at us, we have considered it worth-while to give a fairly extensive survey of this debate and its main arguments. This might seem to be somewhat outside the scope of the present report, but the survey has been written with the direct bearing on the GUME project of the debate in view, even if this is not explicitly pointed out more than once or twice.

When the outline of the Pennsylvania Project, given in the synopsis of the earlier GUME studies (Levin, 1969, p. 6 ff) was written, the results of the two first years' studies (as reported in Smith-Berger, 1968, and Smith-Baranyi, 1968) were available. As a matter of fact, a preliminary report on the third year follow-up was also at hand; however, we then abstained from commenting on more than levels I and II, i.e. the first two years of investigation. Since that time a supplementary report (Smith, 1969a), covering the third and fourth year results as well as complementary statistical treatment of level I and II data, has become available. Various members of the GUME project have also had the privilege of personally obtaining any information desired from Dr Philip Smith, Jr., the project coordinator.

The reader is referred to the above mentioned synopsis for an outline of the Pennsylvania project, its objectives, research design, etc. (Of course we agree with those reviewers of the Pennsylvania project who recommend interested readers to consult the full reports. Any brief critique fails to do justice to the full scope of the findings). The following sketch is for the benefit of readers not acquainted with the Pennsylvania project.

The main purposes were to investigate which of three foreign language strategies was most effective and to determine which of three language laboratory systems was best suited, economically and instructionally, to the development of pronunciation and structural accuracy. The three teaching methods compared were the Traditional Method (TLM), the Functional Skills Method (FSM), and the Functional Skills + Grammar Method (FSG); the three laboratory systems compared were Tape Recorder only (TR), the Audio-Active system (AA), and the Audio-Active-Record system (AAR). The intact school class was the experimental unit. Class assignment was random only across the two functional skills methods (in the case of TLM only teachers who had expressed a preference for that method were assigned to it). The original (= first year's) population consisted of 104 school classes (61 French, 43 German) from nearly as many schools, representing a great geographical variation within the state of Pennsylvania. Of the original 104 classes, 61 remained throughout the second year. After two years, the main finding, obviously not expected by the profession, was that no significant differences existed among strategies on all skills except reading (TLM $>$) as measured by contemporary tests. Nor did the language laboratory of any type, used twice weekly, have any discernible effect on achievement. The criticism that we ventured to pass in our previous report on the research performed thus far (levels I and II) may be summarized thus:

1. The non-random assignment of classes to treatments (in the case of TLM) is a potential source of error in that teacher preference may reflect belief in that strategy, which will breed more enthusiasm for the work and hence increase the chances of better results.

2. The two "Functional Skills" methods do not seem to be very distinct; considering the diffuse difference between FSM and FSG one might suspect that the experiment is, in reality, a comparison between one traditional and one audio-lingual method.

3. No special course material was constructed. The project staff chose five French and four German textbooks out of the twenty-seven which are commonly used and decided which were to be used in each method. Most teachers were thus left with a limited choice. No maximum pensum to be read was established; the different classes could (and did!) cover different amounts of text. Thus

text materials chosen as well as rate of progress in the text-books are possible sources of variation. (As a matter of fact, during the first year, TLM classes covered almost three times as much text as did the FS classes.)

4. An outdated version of the MLA Cooperative Tests (1939-41), apparently favouring TLM classes, was used in one phase of the study.

(A Swedish reader should be aware that the experimental population, compared to Swedish circumstances, was a very select group since only 17-20 % take a foreign language in Pennsylvania; thus even the "low IQ group" would be part of the upper IQ third of the GUME population.)

In the final report (Smith, 1969a)it becomes evident (p. 23) that too few French students remained in the Traditional experimental treatment after three years for meaningful comparisons to be made with Functional Skills classes. The third year summary reads (p. 41): "A sufficient number of German students remained available to the proje staff through Level III to support the conclusions drawn after Levels I and II: 'Traditional' students equaled or significantly exceeded the achievement of 'Functional Skills' students on the MLA Cooperative Classroom Listening and Reading Tests". It should be mentioned that two more conclusions were forwarded, one concerning correlations between measures of teacher proficiency and school class achievement, and one concerning student opinion measures; however, our concern here is with the main results.

Complete data extending over a full four-year period was obtained on 92 students, 72 German and 20 French, i.e. 2 % of the original population. The German students were quite evenly distributed among the three strategies: TLM: 27, FSM: 24, FSG: 21. This sample permitted the computation of an analysis of covariance using the pre-experimental Modern Language Aptitude Test as a covariate. For the French students no such investigation of main effects was possible. The fourth year summary reads as follows (p. 44): "Level IV results support earlier findings that there is no advantage favoring Functional Skills classes in performance on tests designed to measure functional skills. IQ seems to be the best predictor of long-range student foreign language achievement within the secondary school setting". The final report also contains additional information and analyses of the first and second years of study and, most interestingly, a "Condensation of Discussion Conference Proceedings".

The following section is a review of reviews; in the case of the Pennsylvania study, the results of which stirred up emotions and initiated a lot of reviews, this may be a contribution in its own right.

The reviews we shall comment on here are Carroll's (1969) and Wiley's (1969) in the December issue of *Foreign Language Annals*, 1969, and various articles in the now famous October issue of the *Modern Language Journal*, 1969.

In our own review in the previous report (Levin, 1969, p. 6) we stated that the Pennsylvania project would probably become a classic, considering the investment in people and money. Dr. Philip Smith Jr. gives the following factual information on the scope of the investigation (1969c, p. 2): ..... "four thousand two hundred students in one hundred and thirty-two classes representing an investment of three hundred and fifty thousand dollars and over a thousand pages of written materials , ..." Similarly, Carroll says (p. 214): "The Pennsylvania Foreign Language Research Project will undoubtedly go down in the annals of foreign language teaching research as one of the classics. In size, scope, carefulness of experimental design, and importance of results it is unmatched by any previous study of its kind. It has already attracted wide attention because of the apparent discrepancy between its findings and the outcomes that current thinking about foreign language teaching might have led one to expect or to hope for". As the last sentence indicates Carroll is obviously assuming that the profession at large would expect results favouring the audio-lingual methods rather than the traditional. Carroll, although professing that he does not intend to choose sides in the debate, admits his own bias towards a "cognitive code-learning" approach, which undoubtedly has more in common with the TLM than the other two methods in the Pennsylvania study. Perhaps it is this inclination that causes him to take the results, at least to some extent, at their face value (p. 214): "In brief, it (the study) seems to tell us that the 'audio-lingual' emphasis of current FL teaching philosophy is in some way misguided".

Carroll is almost laudatory with respect to the experimental design of the study. "In fact, it is one of the few large-scale studies that has well observed the canons of scientific educational research" (p. 215). This is in agreement with Wiley who states (p. 211): "(In spite of

these criticisms) the design and its implementation were excellent in comparison to other evaluation studies in that no attempt at random assignment of relevant units to treatments is usually made". The following quotation is intended to illustrate the inconsistency between different reviews by qualified researchers (Aleamoni & Spencer, 1969, p. 421): "The study appears to fall more into the category of an *ex post facto* research design while professing to be an experimental design. The *ex post facto* research design does not allow testing for treatment effects but, instead, only permits comparisons between groups, etc., on common variables. In the case of the Pennsylvania Project, data could be collected under this model to determine differences of student achievement in existing but varying classroom conditions, *but the results would not indicate what, if any, effect the classroom conditions had on student achievement*" (italics ours). If this critique were valid, and our own belief is that it is not, the results of the study would be highly suspect.

To return to Carroll, he makes the observation (p. 235) that "the 'Traditional' method used in the study was apparently, in most cases, a 'traditional-modified' method which exposed the student to a considerable amount of spoken language (cf p. 30 below). The most misleading thing about the publicity that has attended the study is the use of the word 'traditional', which will be interpreted by the casual reader as meaning a form of FL instruction that may have been prevalent forty years ago but that hardly has a place in to-day's schools". It is unfortunate that the observation scales used for describing classroom activities were constructed so as not to make control of adherence to method by teachers possible (a fact which has been pointed out by several reviewers); as Carroll observes, TLM students obvicusly used oral language more than they were supposed to (218). If this observation by Carroll is correct, and similarly, if our own statement concerning the diffuse differences between FSM and FSG is correct,then, which *were* the methods being compared in the Pennsylvania project? If we have stressed this point strongly here, it is because we have become aware, during the course of our own work, of the difficulty of keeping the methods distinct (though this must be far more easy in the case of "canned" materials).

Some of the criticisms that Carroll passes on the study are:

Too few classes remain in some of the strategy-system cells for statistical inferences to be made.

The text used, rather than the method, may explain some of the main effects (in Carroll's terminology, the text is a "stowaway variable").

Control of vocabulary load should have been made in the case of the criterion tests.

Sampling of classes was not strictly random.

Some selectivity in the reporting of data can be noticed. ("As this critique demonstrates, the readers of a statistical report sometimes find it necessary to refer to data that the investigators may not think worth reporting", p. 221).

No rationale was given for the choice of covariates.

No two-way analyses of variance were made in order to investigate interaction between strategy and ability.

The tests of "teacher proficiency" were in no sense intended to measure actual *ability to teach a foreign language*; apart from the misleading term, Carroll criticizes the statistical treatment of "teacher data" for being incomplete.

Our review of Carroll's review has been severely selective in that we have hardly made justice to his fundamentally positive attitude to the research completed by the Pennsylvania project staff. Our negative bias has had one aim: to provide the reader and ourselves with a "check-list" when contemplating the present report.

A final quotation from Carroll's review (p. 234): "I do believe that the findings of the study with regard to teaching strategies and laboratory systems are sufficiently solid and replicable to prompt us to rethink methods and objectives in foreign language teaching".

Wiley's review concentrates on the design and the statistical treatment of the results. The most serious defect in the design, according to Wiley, is the non-random assignment of classes to treatments. He points out that the average IQ in schools which had a language laboratory might be different from the IQ in schools without these facilities; thus presence or absence of a language laboratory might be associated with background variables. Because of this possibility it

is unfortunate that no analyses of Pre-test data are reported so that this suggestion could be investigated. "The analysis of covariance may not help in this case since it is sensible to non-random assignment in the presence of fallible covariates as well as to nonlinear regression, where there are large initial differences in the groups" (p. 211). Some other points made by Wiley are: The multivariate test statistics and their associated probability levels are not used. The adjusted means are not reported for the analyses of covariance. Tests of homogeneity of regression do not precede the analyses of covariance.

However, Wiley inclines towards the positive and mentions a number of commendable features of the study, among them "..... the monitoring of the treatment effects which allowed rather more precise definition of the various strategy-laboratory combinations. This is especially useful for those who wish to base decisions on the study" (pp 211 - 212). It is noteworthy that this point, like so many others, has been quite differently commented on by competent reviewers.

In the October issue, 1969, of the *Modern Language Journal*, the Pennsylvania project was fiercely criticized in a number of articles. Some of them were very negative in tone, and one wonders whether the authors had an axe to grind. Anyway, there is reason to believe that at least some objectivity was sacrificed in the heat of argument. We shall be brief in our comments.

Hocking, concentrating on the comparisons between laboratory systems, seems to be accusing the project staff of sabotage as far as the language laboratory side was concerned. Hocking seems to advocate more restricted projects than the Pennsylvania study which he thinks involved too many inponderables and uncontrolled variables. However true this may be, a strong need was obviously felt in the mid-1960's that a study of this dimension should be undertaken.

Clark's main criticisms (p. 388 ff) include: non-random assignment of classes to methods, no clear distinction between methods, faulty scales for controlling teacher adherence to strategy; all these items have appeared above. However, Clark's argument on p. 394 has a strong resemblance with our own discussion of "Hypothetical Treatment Effects" (see p. 22 below): "Within the Pennsylvania Project, the most powerful demonstration of superior pedagogical efficiency for one or another of the three teaching methods would have been for that method to satisfy all of the following conditions: 1) to prove superior for both the

French and German groups rather than for a single group; 2) to show superiority on all three measurement occasions (first- and second-year tests for the original group; first year test for the replication group); 3) to show similar results for closely related tests, as within a single skill area; and 4) to prove superior to both of the other two methods, rather than to only one of these methods. To the extent that these outcomes are not reflected in project results, it becomes necessary to introduce explanatory hypotheses which may become so diverse and complex as to reduce considerably the possibility of identifying a single factor - such as inherent superiority of a particular teaching method - which would account for the observed results". Clark contends that the only safe generalization that can be made for the results of the study is that the majority of comparisons show non-significant differences among the teaching methods. However, he does not accept this as evidence of the pedagogical equivalence of the methods but considers the possibility that true differences may have been concealed by uncontrolled factors.

Otto's review (p. 411 ff) is primarily focused on the area of teacher activities within the project. He contends that the MLA Proficiency Tests do not measure pedagogical proficiency, that several teachers were assigned to teaching strategies against their preference, that assignments were not based on effective screening techniques (which would have helped the project personnel to determine if the teachers had the ability and experience to follow a particular strategy), that the so-called orientation sessions for teachers did not provide exemplary models of effective teaching behaviours for each strategy, that the orientation sessions were no work-shop sessions (which was what was needed), that assistance and supervision was not sufficiently provided, that the *Teacher's Manual* was poorly organized. In short, Otto is strongly negative towards the project, at least those aspects of it which regard the teachers and the part they played.

Valette, in her review (p. 396 ff), mentions one feature which most reviewers have touched on, namely the fact that the complex findings of the Pennsylvania project have been over-simplified and misinterpreted in various press releases. Stressing the disservice such journalism does to both the project personnel and the foreign language teaching profession as a whole, she urges anyone really interested in the results to read the full reports.

One interesting comment by Valette is the following (p. 397): "(Consequently), the section of the Pennsylvania Project which contrasts teaching approaches has almost become out-dated before the results have been disseminated". Her argument is that, in 1969, the distinction between "traditional" and "audio-lingual" is losing some of its relevance because the new traditional texts (the "third generation" texts, in Valette's terminology), make creative use of dialogues and pattern drills whereas (the "second generation") audio-lingual texts give attention to formal grammar. This phenomenon has an obvious resemblance with "the struggle towards the middle", which was discussed in our previous report (Levin, 1969, p. 79).

Some of Valette's criticisms of the study are the same as those discussed above, some may be new: TLM students received more contact with the spoken language than was intended, the contents of the Cooperative tests favoured TLM students (TLM students did much more poorly on this test, however, than one would have anticipated), the criterion test was too difficult, the student opinion scale is dubious (an expert on attitude testing ought to have evaluated the instrument), etc.

Her main point on the use of the language laboratory is that, in the lab, one tape was played to the entire class; thus the lab was not used for individualization. "..... we must distinguish between the physical installation which we term a language laboratory and the use we make of that laboratory" (p. 404).

Finally, mention should be made of Valette's proposition that, in modern languages, criterion-referenced tests should be developed. According to her, the Pennsylvania project had specified "expected levels of proficiency" but had no tests available to assess whether the pupils reached those levels.

The last review in the "October issue, 1969" that we shall comment on is that of Aleamoni and Spencer (p. 421 ff), who are very critical: "In general, the objectives of the study are stated more broadly than the study seems capable of handling; and it covers areas so diverse that it would be difficult for any study to accomplish them" (422).

The authors criticize the project for being unwieldy and unmanageable.

Furthermore, the project staff is accused of being subjective and biassed in planning the study: "Many of the statements in the early

pages of the reports are statements of belief, opinion, or attitude, which set the stage for the research design. These statements appear in the reports without evidence or documentation " (p. 423). Some of the more specific criticisms concern the (alleged) misuse of the interest, attitude, motivation and teacher factor scales, the decision not to include students for whom complete data were not available, use of the same test as both a covariate and a criterion when the covariate had been subject to the effects of the treatment, etc. Of all the recommendations to the teaching profession, forwarded by the project staff at the end of the reports, none seem to escape Aleamoni's and Spencer's criticism.

Later on Dr. Smith wrote a reply to the October, 1969, *Modern Language Journal* (Smith, 1969 c). When he states that "Some reactions have been of the highest professional quality, some reflect simply a lack of understanding, others smack of panic" (p. 3), he refers to all reviews until that date. Concerning the specific MLJ reviews he contends that they "often present a distorted view of the Pennsylvania Studies in that they suffer from (1) a narrow and insulated viewpoint; (2) overt hindsight; (3) personal interpretation; (4) inconsistency; and (5) obvious oversight. This is tragic, especially in that the *Modern Language Journal* attempts to be a responsible professional journal but will not protect its contributors nor its readers from obvious oversight, choosing to let errors stand as definitive statements of the research" (pp.5-6). For some reason, the reviewers had had no contact with the project staff, which might have led to a correction of errors - if there were such - or at least to a relaxed atmosphere, more advantageous to scientific cooperation.

Dr. Smith points at a number of issues where the reviewers have different, not to say opposed, opinions. However, we shall not discuss his counter-arguments here, nor try to pass any kind of value judgment on them. It seems a difficult task to make a reliable and comprehensive evaluation of the Pennsylvania project in all its complexity. At any rate, the contrasting views of competent researchers on various aspects of the project, is one indication of this.

Whatever significance the project results will have in the long run, the following statement may be made with confidence: being contrary to the expectations of many foreign language teachers, the project results have initiated a debate that will in turn initiate wholesome rethinking on various aspects of foreign language teaching methodology.

EXPERIMENTATION IN A FIELD SETTING - SOME REFLEXIONS

Comparative Experiments - Pros and Cons.

The present study is a case of variable-manipulating, comparative experimentation in a field setting. Since the general value of such research has occasionally been questioned, a comment may be appropriate.

A classic in this debate is Scriven's (1968) article, where the principles of formative and summative evaluation are introduced and, which is of greater interest here, where Cronbach's (1963) "despair over comparative studies" is optimistically contradicted. "If we have really satisfied ourselves that we are using good tests of the main criterion variable (and we surely can manage that, with care) then to discover parity of performance *is* to have discovered something extremely informative. 'No difference' is not 'no knowledge'" (Scriven, p. 67). Scriven apparently holds the view that the comparative field study has a definite (though by no means unlimited) place in evaluation.

A representative of the negative attitude towards field experimentation is Grittner (1968) who, when commenting on the bulk of studies presented by Stephens (1967), concludes: "In short, half a century of such 'research' has told us almost nothing about the relative superiority of one educational strategy over another!" (Examples of the areas which Stephens reported on are the following: large vs. small schools; large vs. small class size; accredited vs. non-accredited teachers; progressive vs. traditional education; live teachers vs. TV; lecture method vs. discussion method; team teaching vs. traditional teaching; and homogeneous vs. heterogeneous grouping of students). "Tables showing standard deviations, covariance, F-ratios and the like are very impressive; however, if the ultimate result of such studies is that they cancel one another out, perhaps we should ask for a cease fire while we search for a more productive means of investigation" (p. 7).

Wiley (1969) makes a distinction between conclusion- and decision-oriented research. The former is performed so that the investigator may draw conclusions about the phenomenon he is studying. Conclusions,

however, are tentative by nature and may be modified as more evidence is accumulated. Decision-oriented research, on the other hand, is performed to gather evidence which will be used for generating decisions about actions to be taken. Wiley gives the example of a school superintendent who cannot wait for twenty-five years of accumulated evidence before deciding whether to purchase a language laboratory. If he does so, he will really have decided against it (p. 209). Wiley further argues that the concern for the quality of evidence must be greater in the case of decision-oriented research; decision-makers cannot wait for ambiguities to be clarified by subsequent investigations. Under these circumstances, the methodology of research becomes extraordinarily important.

The point that we want to make here is that Wiley seems to come rather close to the traditional design proposed by Campbell and Stanley (in Gage, 1963) when suggesting proper evaluation methodology. The main difference seems to be Wiley's greater concern with the criterion tests to be used in program evaluation ("It is not individuals among whom we wish to discriminate; rather it is programs", p. 208). His philosophy of evaluation thus seems to be quite similar to Scriven's. In spite of the difficulty of constructing reliable evaluation instruments, Wiley seems to be in favour of experimentation in school settings.

Stanley (1970) regrets the present state of affairs in educational research, which, according to her, is characterized by the paucity of controlled experimentation. "Apparently there is more lack of intent, money and technical resources than of available, applicable methodology. Those critics of experimentation for evaluation who say that controlled, variable-manipulating experimentation may be splendi for stands of alfalfa and weights of pigsbut inapplicable to educatio do not adequately appreciate the generality of Fisherian and neo-Fisherian methods. ....Inflexibility is more in the minds of plann researchers, and critics than in the methodology itself. Of course, there is no royal road to new knowledge; it is not easy to experiment with human beings, whether they are medical patients or school pupils In my opinion, however, controlled experimentation and some quasi-experimental designs are important methodological tools of the educa tion evaluator. Recent attempts to rule experimentation inapplicable because other methods are also useful seem misguided" (p. 107).

The survey of opinions for and against experimentation in the natural school setting might have been made more extensive. For the moment, however, we shall be content with this list of contrastive views. Textbook writers in the branch of educational research often present an almost overwhelming list of difficulties of experimentation but end up with words of encouragement, urging the student to use experimental methods whenever they are feasible.

Later in this report we shall return to the question of comparative studies and their value as a research activity. However, let us conclude this section by quoting Wiley once more (ibid, p. 210): "In any research study, especially one conducted in a field setting, it is impossible to do everything 'right.' There are always going to be unanticipated contingencies and contingencies which, although anticipated, are practically (usually monetarily or cooperationally) impossible to avoid. The main goal is to spend the most time, effort, and money to avoid the most 'important' pitfalls to the validity of the findings and their interpretation. One problem is that the 'importance' or relevance of each pitfall is different for different individuals".

The GUME Project - Some Comments.

In one of the earlier GUME reports (Levin, 1969, p. 27 ff) our first three studies were discussed in relation to Carroll's chapter "Research on Teaching Foreign Languages" in Gage's Handbook (Gage, 1963, p. 1060 ff). Here we shall avoid unnecessary repetition; however, a few points will be made.

In GUME 4, as in the first three projects, we do not have the advantage of what Carroll calls a natural zero-point in second-language acquisition. The experimental population consists of pupils in their third year of English, as compared to the fourth year in the previous studies. As a matter of fact, the GUME 4 study was performed during the spring term whereas the earlier studies were performed during the autumn term (with one exception: GUME 3 in January); thus the real difference between the studies with respect to general competence was probably a small one. Although prior knowledge in English is controlled statistically by analysis of covariance (to the extent that our Achievement test measures this), it is obvious that the *amount* of treatment (teaching) must be large for *differences* between the various treatments to appear. We said earlier in this report (see p. 3) that our

three teaching methods (and certain other factors) were modified so as to increase the probability of revealing true method differences. However, our strategy of making the three methods "optimal" may have worked the other way round, thus reducing (artificial) differences between the methods. The research problem, in a nutshell, is then: Should one use radically different treatments, thereby increasing the chances for a "posivite" outcome but decreasing the external validity of the findings, or should one construct different but "realistic" methods that might be used later in school, thereby decreasing the probability of obtaining "positive" results? Posing the problem in this manner is perhaps somewhat naïve, but it has to be solved, anyway. In GUME 4 we have decided to pursue the latter course for two main reasons. Our three methods have the theoretical psychological background formulated by Carroll (1965, p. 101); they are thus not *ad hoc* creations to form contrasts in an experiment. Secondly, the debate on methods in language teaching in Sweden (see p. 30f below) has created a kind of polarization which we wanted to shed some light on. We considered it more worth-while to test realistic methods at the risk of not obtaining positive results, than to try to get such results and then be left with the question how to interpret these results and what use they can be put to.

Another circumstance decreasing the probability of obtaining positive results is the fact, not particular to GUME but rather general, that pupils vary in a number of aspects, and that *this variation is treated as error in the analyses*. Incidentally Carroll (1969, pp 233-34), when reviewing the Pennsylvania Study, notes that "another unassailable fact arising from the study - and one that carries at least some surprise - is that *classes* vary enormously in average performance". Without anticipating our results we may perhaps state that the same observation was made in the present study; the differences between the school classes, let alone between the individual pupils, was enormous. Hopefully a good deal of this variation is held constant in the analyses of covariance, but it would be a false assumption to believe that all that variation, for instance in Post-test scores (an indication of a corresponding variation in general ability, motivation, reading facilities in the home, day-dreaming tendencies and what not) could ever be held constant, experimentally or statistically.

Hypothetical Treatment Effects. x)

The present investigation implies a comparison between three teaching strategies. No assumptions are made about the superiority of any one method; to use a different terminology, the null hypothesis is being tested. The experimental design should be such as to make interpretations of the results as clearcut as possible. Of all the theoretically possible outcomes, some are more difficult to interpret than others. In this section we will briefly discuss specific interpretation problems that may arise.

The three teaching strategies being compared are

Im Ee Es

On the one hand the effect of explanations is compared with the effect of non-explanations, on the other one method utilizing the source language (Swedish) is compared with two methods utilizing the target language (English). An ideal design for isolating the effects of explanations/nonexplanations, source language/target language would have to include an Im_s, i.e. Im-Swedish, variant, however, since such a method is impossible per definition, and, accordingly, could not be included in the design, the interpretation problems indicated above will arise in certain cases.

When comparing three strategies, the following main results are possible:

a) two methods equal and better than the third (3 possibilities)
b) one method better than the two others, they being equal (3 possibilities)
c) method X better than method Y better than method Z (6 possibilities)
d) the three methods equal.

According to a) above, the following three outcomes are possible in the GUME project:

1. Ee = Es $>$ Im
2. Im = Ee $>$ Es
3. Im = Es $>$ Ee (?)

In case 1. the facilitative learning effect is unequivocally due to the explanations, in case 2 to the use of English, whereas in case 3 the result could not be logically explained. The superiority of methods Im and Es can be accounted for neither by reference to language of instruction nor by explanations.

---

x) This section is identical with the one in Levin (1969, p. 29 ff).

Correspondingly there are three possible outcomes according to b) above.

4. Im > Ee = Es
5. Es > Ee = Im
6. Ee > Im = Es (?)

In case 4 the non-explanation method is unequivocally better than the two explanation methods, in case 5 the facilitative effect can be traced to the use of the source language, whereas in case 6 the outcome is impossible to interpret. According to c) above, six results, approximately identical to the six just presented, are theoretically possible. Our intentation here is only to predict difficulties of interpretation in general, and we will not discuss interpretation problems under c) further. Concerning d)(the three methods equal) it should be remembered that such an outcome does not *prove* that there exist no differences between the methods (as is well known it is a logical impossibility to prove the null hypothesis). One possible explanation might be that the experiment, as it was planned and executed, did not succeed in detecting actually existing differences between the methods.

To sum up:
The experiment makes possible comparisons between three methods of instruction. Theoretically thirteen different outcomes are possible. Some of them would be impossible to explain, or rather, would arouse doubts about the experiment, notably the experimental control of the three teaching strategies. We may have good reason for returning to the interpretation problem in the results section.

METHODS IN FOREIGN LANGUAGE TEACHING

Introduction

One important aspect in the planning and reporting of a comparative study like the present one is the exact definition of the different methods used in the project. Even the Pennsylvania Project, which has been discussed at some length above, seems to have failed to a certain extent in this respect as many of its critics point out, among them Carroll, Clark, and Valette (see pp. 9, 14, 16 above). In the GUME project three different methods were compared. In studying and interpreting the results it is important that the reader has a clear picture of what is compared, in what respects the methods differed and which were the points of comparison. As a background to the GUME methods a short survey will be given of some of the ordinarily used terms.

Some Well-known Methods.

How many different methods do foreign language teachers have to choose between and in what respects do they differ and what are their characteristics? These questions are more difficult to answer than one might think.

Mackey (p. 151), after an historical survey, lists no less than fifteen different methods and gives short characteristics of them. Titone (p. 97) uses three main headings, the formal, the functional, and the integrated approach, and then subdivides the second of these into five different methods. Carroll (1966, p. 101) has tried to arrange all competing methods in two groups, based on two opposing psycholinguistic theories, the *audio-lingual habit formation* theory and the *cognitive code-learning* theory. Rivers (1968, p. 11) seems to have a similar classification in mind when she groups the various methods into the categories *activists* and *formalists*.

One reason for this seemingly chaotic state of affairs might be that language teaching is such a many-faceted art. How should vocabulary be taught? How grammar? In what respects should elementary school English (as a foreign language) differ from high school and college English? To what extent can/should/must the linguistic differences between English and Russian effect methods? Teachers who agree on one point may very well

differ on another. And how can differences between methods best be described?

Mackey has constructed a "method profile" (see e.g. pp. 317-318), which however seems fairly difficult to read.

One might simplify matters and arrange methods along a continuum (see fig. 2), putting an extreme grammar-translation method at one end and an extreme direct method at the other. It would then probably

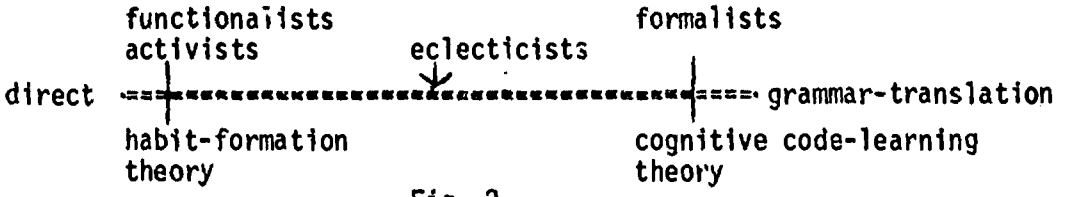


Fig. 2

be possible to divide this line into three parts: the two extremes which differ radically from each other and from most of the in-between gradations, and the largest part along which the methods used by most language teachers would in all likelihood be arranged. This corresponds well to Casey's "Methods Profile" developed in an experiment concerning the teaching of English in some Finnish schools and being an attempt at quantification of method (Casey, 1968, p. 6).

Most advocates of a formalist kind of teaching would be somewhere to the right of the middle, including those who favour translation, theoretical grammar and a lot of written work. Towards the left would be activists with the direct method proponents close to the dividing line (it is probable that the Berlitz method, the well-known American private school, would be beyond that line at the extreme end), followed by audio-lingualists like Brooks. At the centre of the line we would find the eclectic method (Rivers, 1968, p. 21), or perhaps rather the eclectic methods. It is probably correct to characterize Wilga Rivers as the most outstanding eclecticist and her two books as the most authoritative formulation of this middle-of-the-road method.

The Authorized Curriculum for Swedish Schools.

The official curriculum for all Swedish schools on the compulsory level (Läroplan för grundskolan) sets down both goals and recommended methods for the teaching of English and the second foreign language (French or

German). The versions of 1962 and 1969 (Lgr 62 and Lgr 69) differ to a certain extent, but it would probably be correct to characterize them both as proponents for an eclectic method even if they represent a position rather to the left of the middle and with the latest version (Lgr 69) to the left of the older one.This means that the audio-lingual kind of structure drill method for teaching grammar is proposed and that teachers are advised to be restrictive in the use of theoretical grammatical explanations. Also, a direct method kind of monolingual method for teaching vocabulary is advocated rather than a translation method. It does *not* mean, however, that the teacher is forbidden to use grammatical terminology or translation when this must be judged the best method under certain circumstances.

GUME Methods

Within the GUME project we have chosen to use the terms Implicit and Explicit method rather than any of the accepted terms like direct method and grammar-translation method. There are two main reasons for this:
1. The established terms are unclear and filled with connotations, good or bad as the case may be.
2. The project has not investigated the teaching of English as a foreign language, not even the teaching of English on a certain level. It has tried to investigate *the teaching of grammatical structures* in English on a certain level. The established terms normally refer to the teaching of a foreign language in general. In the project we have not investigated and not expressed any opinions on how vocabulary should be taught, when the student should be introduced to the written language, etc. The three methods used in the project and described in some detail below, are thus not on a par with other names of language teaching methods discussed above.

It should also be noted at the outset that we have not tried to investigate methods as different as possible, represented by the extreme ends of the line in fig. 2 above, but rather methods which are all in the central part of the continuum and which would all find proponents among language teachers. They could all be said to fit into the Swedish curriculum, since this is written in such a way as to allow a wide variety of methods and procedures to suit different teachers and pupils.

It should also be stressed that our methods are not just *ad hoc* creations to make up a nice experimental design with contrasting methods, but rather they are an attempt at putting the two theories formulated

by Carroll (1966, p. 101) to the test, one a pure habit formation theory, the other a cognitive theory, which does not eliminate practice and which does not necessarily mean a deductive method with rote learning of rules etc. as some grammar-translation methods would have it. A more comprehensive description of the three methods will be given in the next chapter where the lessons are described in some detail. Here we shall try to give some of the theoretical considerations behind the three strategies.

Implicit. The implicit method corresponds to Carroll's habit formation theory, based largely on Skinner's experiments and writings. It is well in line with a "pure" audio-lingual method as it has been described by, for example, Nelson Brooks (1960, p. 47): "The single paramount fact about language learning is that it concerns, not problem-solving, but the formation and performance of habits." Brooks, however, does not forbid the giving of generalizations after a grammatical structure has been practised. But "pattern practice" or "structure drill" which "makes no pretense of being communication" is the corner-stone of this method. This is also in keeping with the recommendations in Läroplan för grundskolan, Supplement, Engelska (1969), where it is stressed (pp. 12-14) that "The learning of grammatical phenomena takes place through systematic practice", and: the exercise should be presented to the pupils "in such a way that the pupils understand what the teacher wants them to do". "The insight into the build-up of the language, which the pupils are supposed to arrive at, is achieved mainly through systematic practice". Generalizations should come in late and preferably be formulated by the pupils themselves which proves "that the pupils have reached insight *through* the exercise". Wilga Rivers (1968, p. 43) points out that in some materials, especially for junior high schools, "these generalizations are omitted because it is believed that the very design of the materials will lead to an inductive apprehension of structural relationships". This is, according to Rivers (p. 48) typical of the direct method, where the student "must acquire the meanings of words and the functioning of structural patterns inductively with very few props to help him", and she feels that this makes it particularly difficult for the less gifted pupils.

Our Implicit method is thus an inductive method in which the pupil is left to draw what conclusions he can from drills, very carefully structured drills, and it is our belief that this method is used in many classrooms today.

The Explicit Methods. Both our explicit method would fall under Carroll's category Cognitive code-learning theory which stresses the intellectual (cognitive) understanding of what one is doing. This is not an old grammar-translation method since a large part of the time is taken up by structure drills, the same as in the Implicit method. Carroll pointed out in 1965 (MLJ p. 281) that the audio-lingual approach, no longer "in step with the state of psychological thinking", was "ripe for a major revision, particularly in the direction of joining with it some of the better elements of the cognitive code-learning theory". This mixed method would fairly well correspond to what Rivers (1968, p. 21) has called the eclectic method: "The true eclecticist, as distinguished from the drifter who adopts new techniques cumulatively and purposelessly, seeks the balanced development of all four skills at all stages". This is roughly the kind of technique proposed by Palmer (1921), and the method recommended by Rivers, a modified audio-lingual approach, would also fall into this category.

In the Explicit methods the pupils were given generalizations (which is probably a better term than the one we used, explanations) about what they were doing in the drills. This is in line with the normal audio-lingual approach, as Rivers (1968, p. 43) points out. She expresses the opinion that in drills based on uncomplicated structures the students can "establish for themselves what the point at issue is, and little or no explanation is necessary" but with more difficult structures which form a contrast with the native language "the teacher should make sure that the students understand what they are expected to learn by the drill" (p. 82). This is an excellent statement of one of the points at issue in our project, and what we have tried to do is to establish where to draw this line.

The *Explicit-English* method is so far from being in line with the traditional approach that it could rather be characterized as a direct method, which "gave structural explanations as well as exercises in the language" (Rivers, 1968, p. 84). This is in line with what the Läroplan för grundskolan, Supplement, Engelska, p. 14 prescribes: "Every grammatical rule must (sic!) be formulated with English as the starting-point". The writer of these recommendations also *knows* that if "some-any" are translated "this will give rise to a mixing of them which might be avoided" if they were practised separately, which will make confusion "impossible since the two words, in a given context, exclude each other". This is a point which we wanted to investigate.

The *Explicit-Swedish* method is probably the most widely used. It is advocated by many teachers and it has been recommended by, for example, Professor Alvar Ellegård in various newspaper articles (DN 3/1, 8/2 1969, cf Edwardsson, 1970), with the assumption that new developments in linguistics and psychology have overruled the tenets behind the "New Key" movement (i.e. the audio-lingual method in post-world-war-two USA). The Official Curricula for Swedish Schools (Läroplan för grundskolan) which does not forbid the giving of explanations or even rules in Swedish has been understood to do so, and this has given rise to some debate in which a return in the direction of what might be termed an explicit-Swedish method has been advocated. Whether this method is best suited for weak students, as Rivers (1968, p. 85) presupposes, is one of the main objectives of the project to investigate.

To sum up:
the Implicit method corresponds to a "pure" audio-lingual method without generalizations,

the Explicit-English method corresponds to an audio-lingual method with direct-method generalizations in the target language,

the Explicit-Swedish method corresponds to an audio-lingual method with explanations or generalizations and comparisons with Swedish structures.

It may be of interest to compare the GUME methods to those of the Pennsylvania Project. An attempt at visualizing this is made in fig. 3. The explicit methods compare roughly to the Functional Skills Grammar method, an audio-lingual method including grammatical generalizations. The Explicit-Swedish method is perhaps a little more "traditional" than the Explicit-English. The "pure" audio-lingual method is called Implicit and Functional Skills in the two projects respectively. The Implicit is perhaps a little further to the left than the FSM since in this method grammar is not totally forbidden (as it was in the Im method); cf Smith-Berger (1968, p. 21), criteria for FSM: "D. Grammar - 1. Descriptive rather than prescriptive, 2. Incidental to functional skills being taught".

As for the position of the Traditional method in the Pennsylvania study there seems to be some debate. The *term* has been severely criticized, and Carroll (1969, p. 219 and *passim*) points out that this method might have corresponded to his own suggested "Traditional-Modified" because

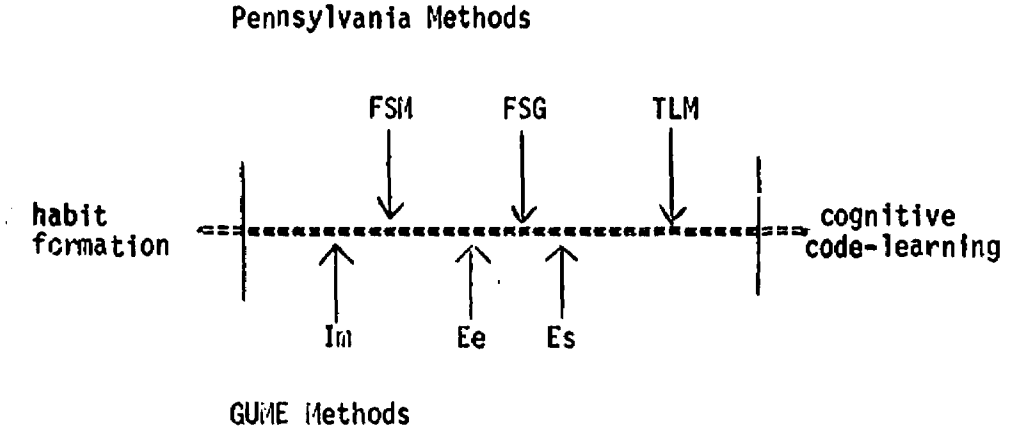


Fig. 3

of the amount of foreign language used in class, the use of tape-recorders etc. He also says that "the most misleading thing about the publicity that has attended the study is the use of the word 'traditional'" (p. 235).

To sum up:

The differences between the methods compared were somewhat larger in the Pennsylvania study than in GUME, but in neither case were they as large as is theoretically possible. The methods were all of a "middle-of-the-road" kind as practised in classrooms throughout the world today.

Current Debate.

The debate on language teaching problems has been extremely vivid in Sweden in the last few years. A brief review of some of the discussion was given in our first report (Lindblad, 1969, pp. 27-28). Quite recently most of the articles have been brought together and commented on in a book by Roland Edwardsson, "Språkdebatten 1969 - 1970" ("The Language Teaching Debate 1969 - 1970").

What is of interest to our project are those arguments which deal with the teaching of grammar. The "action of the 2000", a long letter from 2000 teachers in the Swedish 'gymnasium' (sixth form) to the Minister of Education, demanded the acceptance of grammar books written in Swedish, making comparisons with and giving rules in the mother tongue. The method recommended would probably come close to our Es

method. This action concerned pupils aged 16-19, however.

Professor Alvar Ellegård in Gothenburg, the originator and sponsor of the GUME project, who had started the debate in 1969 by proposing a re-thinking in methods considering new findings in linguistics and psycholinguistics and in educational research (mainly the Pennsylvania Project), started a new debate with an article in June, 1970: "Dåliga språkkunskaper är direktmetodens fel" ("Bad language proficiency is the fault of the direct method."). In this article and in the ensuing debate, primarily with Per-Olof Hensjö, Ellegård suggested a concentration on vocabulary learning at lower levels (in 'grundskolan', the 9-year compulsory comprehensive school, pupils aged 7-16). He proposed an exclusion of grammar both in the form of structure drills (as the audio-lingualists would have it) and of theoretical grammar (as the formalists suggest). This might be an unfairness to the brighter pupils, but Ellegård wanted to take this risk for the sake of the non-streamed comprehensive school and he felt that these pupils would easily make up for this loss at higher levels, in 'gymnasiet'. The last two articles by Hensjö and Ellegård were called "Skall grammatiken kastas ut?" and "Nej, men drillövningarna" ("Are we to throw away grammar?" - "No, but structure drills"). The end of this discussion was that Hensjö, who is a defender of the direct method, stood up in defence of grammar (taught by drilling and not rules, of course) and linguistic strictness.

Edwardsson, who has criticized what he feels to be an undue looseness in the policy of the National Board of Education, has also advocated a strictness and demand for correctness in grammatical matters, and in his comments on the above discussion he sides with Hensjö. This is an indication of the complexity of the discussion. Those who are, by newspapers and the public at large, taken to be on opposite sides in the debate often agree, and vice versa. If the various arguments were plotted along the continuum introduced in figures 2 and 3 above (pp.25 and 30), a larger amount of clarification might be won.

It is in this setting of uncertainty and opposing claims that we have, in the project, tried to shed some light on the problem of the place of theoretical grammar in teaching grammatical phenomena. We have of course made all possible effort not to favour any of the methods used in the project. The lessons we made were the best we could produce within the framework of the study.

A DESCRIPTION OF THE LESSON SERIES

Schedule

The teaching phase of the project consisted of 12 40-minute lessons. Three lessons were given every week, and the project thus took four weeks, exclusive of testing time. All lessons were pre-recorded and had an actual running time between 32 and 38 minutes. Twelve booklets of 7-9 pages, one for each lesson, were prepared. They contained reading texts, tables and other background material for drills, pictures and written exercises. The teachers handed out the booklets and started the tape-recorder, and then their sole - but important - function was to supervise the pupils and see to it that they worked properly. Especially in connection with the oral drills the teachers had to make the pupils answer; they indicated individual students who were supposed to answer and activited the pupils in repeating after the tape. The teachers were not supposed to give any help of a linguistic kind.

In preparing the material, we always made the implicit lessons first. They were the backbone of all lessons. Then all explanations for the explicit groups were written and timed, and finally certain exercises or parts of exercises in the implicit lessons were replaced by these explanations. Great attention was paid to the length of the lessons so that all three methods should get exactly the same amount of teaching time. The final figures for this are given in table 1.

Contents

The following grammatical phenomena were practised: the s-form of the verb in the third person singular present (he gets up late); the present and past continuous tenses in contrast to the simple present and past (he is playing the piano - he plays the violin, she was reading when he came in); preposition followed by an ing-form of the verb (he is good at dancing); the position of adverbs of time (he is always late, he always comes home late); the some-any dichotomy, including something, somebody, anything, anybody; the do-construction in questions and negative sentences, both in the present and the past tenses, and in all persons (does he like tea? - yes, he likes tea very much etc.); and finally the regular past tense in -ed (he walked home).

Table 1. Total Running Time of the Twelve Lessons in the Three Methods.

Lesson	Minutes per lesson Im	Ee	Es
1	37	37	38
2	37	38	37
3	36.5	36	37
4	37	36.5	37.5
5	36.5	36	35.5
6	35.5	36	36.5
7	31.5	32	31.5
8	32	32.5	32
9	36.5	36.5	36
10	36	36	36
11	36.5	36.5	36
12	32	32	32.5
Total	424	425	425.5

Ee and Es had almost the same total running time, Im had 1 minute less than Ee and 1.5 less than Es.

The distribution of these various grammatical points is shown in table 2 , where we indicate in which lessons these things were actually practised (not just occurred).

An attempt was made to vary the lessons as much as possible. Many different activities alternated: listening, oral drills with different stimuli, written exercises and reading. All four language skills (listening, speaking, reading, writing) were practised, but the main objective was the learning of the above-mentioned grammatical structures and the pupils' ability to use them; listening and reading, the passive

Table 2 . Grammatical Structures and When They Were Practised.

Structures	Lessons											
	1	2	3	4	5	6	7	8	9	10	11	12
the s-form	x	x		x	x	x	x	x				x
present continuous						x	x	x				x
past continuous											x	x
prep + ing-form										x	x	x
adverbs of time	x	x		x						x		
some-any		x	x	x	x	x	x	x	x	x		
do-questions	x	x	x	x					x			
does-questions	x	x		x	x	x	x	x				
don't									x	x		
doesn't									x		x	x
did-questions									x	x	x	
past tense									x	x	x	

skills, were thus of secondary importance and in speaking no kind of pronunciation control was introduced, and vocabulary learning did not occur except incidentally. Although the lessons outwardly resembled ordinary lessons in that they were varied and included practice in all four skills, they differed in that the goal was more limited; compare p.26 above where reasons for the new names for our three methods are discussed.

One Lesson Described

It would take up too much space to describe all twelve lessons in detail. Only by listening to the tapes with the pupil booklets in front of himself, can one get a full picture of what the lessons were like. As an example, one lesson, lesson 7, will be described here in some detail.

First the pupils listened to chapter 3 of a story which continued through five lessons and which contained a large number of examples of 'some' and 'any' and their compounds. The pupils had the text, one page,

in front of them. A few questions were then asked on the text and the answers, most of which contained examples of 'some' or 'any' were given; the pupils were just listening. This first part, during which the pupils were silent (but hopefully not completely passive!) took just over 4 minutes.

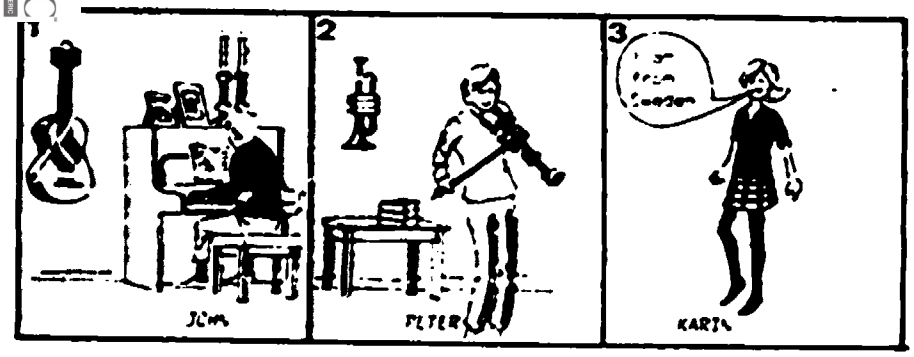
Then the pupils were asked to turn to page 2 (see fig. 4 for a diminished copy of it). This is a mechanical drill of 'not anything' in the sense of 'nothing'. First the pupils listened to the whole dialogue and then they were asked to take over Bill's part. Normally drills of this kind were made as 4-phase: Tom's sentence is the stimulus, one pupil speaks Bill's part (the teacher points to a pupil who answers), the tape gives the right sentence, and then the whole class repeats this. Working with this page took about three minutes.

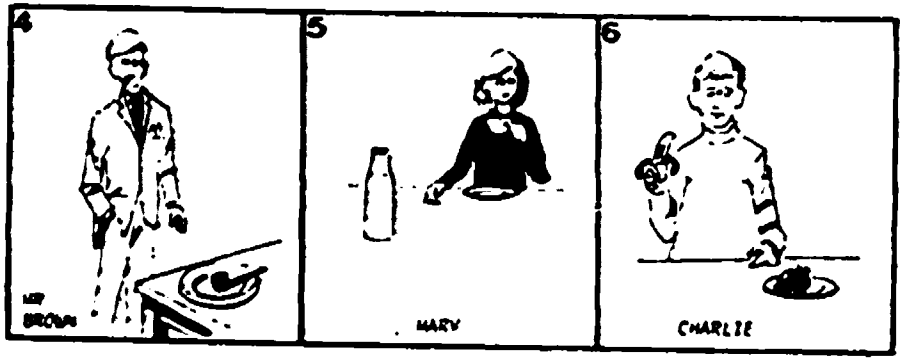
After this they were allowed to relax while they listened to a song, the text of which was given on page 3 of their booklets.

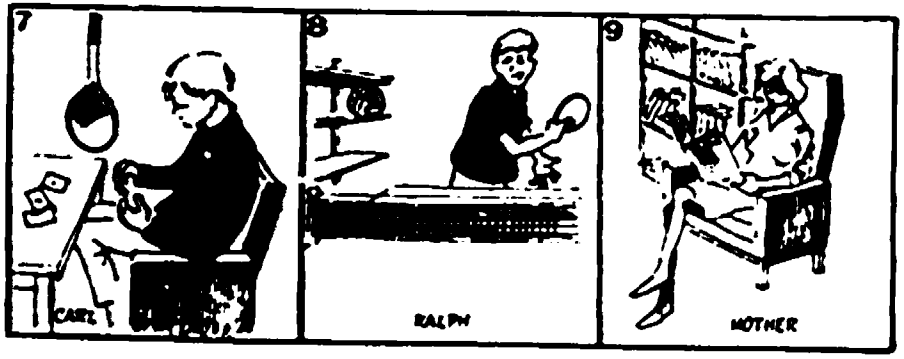
On page 4 the pupils practised 'any' in questions in a written drill. After a short introduction in Swedish they were given 4.5 minutes to write in. The teacher had an overhead copy of the page with the correct phrases in it. He put this on the overhead projector after 2 minutes, so that the pupils could correct what they had written as they got ready. The weakest pupils who might not have known what to write could copy the correct phrases, but experience showed that very few did that. When one minute remained soft piano music was played on the tape to warn the pupils that it was time to start correcting what they had written. Not all of them had time to write everything.

Next the pupils looked up the pictures on page 5 (see fig. 4). In all these pictures there is somebody doing something at the moment, but there is also something to indicate that at other times he or she does something else, e.g. in number 1 John is playing the piano, but on the wall is his guitar: "He plays the guitar very well". This is meant to practise the meaning of the simple present and the present continuous.

First the pupils listened while the voices on the tape spoke about the pictures, next they were asked to repeat after the tape, and then they answered questions,like "Does John play the guitar?", "Is he playing the guitar now?", "What is he playing?": for Swedish pupils, in whose language the difference between the simple and continuous tenses does not exist, the difference in meaning poses a greater problem than the forms. This exercise took a little over 12 minutes in all.







A Dull Day.

Tom and Bill come home. Life is dull. There's nothing they can do.

Tom: There's nothing on TV tonight.
Bill: There isn't anything on the radio either.
Tom: There's _____ on the table.
Bill: There isn't _____ in the fridge either.
Tom: _____ to drink.
Bill: _____ to eat either.
Tom: _____ to smoke.
Bill: _____ to drink either.
Tom: _____ to read.
Bill: _____ to listen to either.
Tom: There's nothing to do.
Bill: But we must do something.

Fig. 4: Two Pages from the Pupils' Booklet (diminished)

Finally they had pictures 4, 7 and 9 reproduced on page 6 in their booklets and were asked to write down answers to questions similar to those that they had answered orally before. They had 4 minutes to do this. They had an overhead key and music to warn them that time was up just as in the previous written exercise.

The total running time of this lessons was 31.5 minutes; this happens to be the shortest lesson of all.

Explicit Lessons

The comments given in the explicit groups were sometimes very short, like "When you write this, remember to have the 's' after 'he', but not after 'I' and not after 'they'", sometimes very long, taking 4 or 5 minutes. In the latter case they were combined with written or oral practice, they were not just long lectures on theoretical grammar but rather commented drills where the pupil was "taken by the hand". No pre-determined fixed time of explanations per lesson existed, as it did in our previous experiments (the exact time for the explanations is given in table 3). The explanations were meant to be "optimal", simply defined as the best we could produce for our purposes and taking as long as they had to. The explanations in Ee and Es were of almost equal length, however, even though this was not a fixed condition. There were between two and eight explanations in each lesson.

In GUME 1 and 2, as a contrast, we tried to work with a stricter, theoretical plan: the explanations there took up about 30 % of the time, consisting of three 3-minute comments per lesson (which sometimes led to an unnatural "stuffing"), and the formulation of the explanations had to follow a strict pre-determined plan: in GUME 1 a kind of transformational approach was attempted, in GUME 2 we kept to a strongly semantical kind of explanation, making very little of the various structural surroundings of some-any. In GUME 4 we used all kinds of explanations, whether they should be termed traditional, structural or transformational.

The most common procedure in GUME 4 was to have a short introduction either in the form of a few examples that the pupils just listened to, or in the form of a short drill, then came the explanation, and after that followed the main body of the drill. This seems to be slightly different from the common audio-lingual practice: "(the) generalization sets out in organized form what he *has been doing* in the drill" (Rivers,

Table 3: Exact Times for Explanations in the E Groups

Lesson	Ee	Es
1	1'02"	1'17"
	2'20"	2'35"
	57"	53"
	23"	34"
	35"	38"
	1'00"	1'05"
Total	6'17"	7'02"
2	26"	26"
	1'17"	1'20"
	31"	42"
	15"	10"
	37"	42"
	1'08"	44"
	1'20"	36"
	40"	35"
Total	6'14"	5'15"
3	10"	10"
	1'10"	2'15"
	7'20"	6'55"
	1'25"	1'50"
	1'37"	1'20"
Total	11'42"	12'30"
4	3'06"	3'04"
	2'27"	2'37"
	17"	24"
	1'27"	1'20"
	2'03"	2'37"
	2'12"	2'38"
Total	11'32"	12'40"
5	50"	51"
	4'30"	4'17"
	50"	52"
	2'47"	2'13"
	32"	32"
Total	9'29"	8'45"
6	1'28"	1'53"
	26"	25"
	25"	31"
	54"	56"
Total	3'13"	3'45"
7	25"	26"
	39"	42"
	3'13"	3'05"
	42"	24"
Total	4'59"	4'37"
8	1'33"	1'41"
	4'27"	4'01"
Total	6'00"	5'42"
9	48"	53"
	1'02"	39"
	19"	18"
	2'47"	2'32"
	2'06"	2'25"
	59"	52"
Total	8'01"	7'39"
10	48"	45"
	2'00"	1'56"
	2'10"	1'57"
	1 53"	2'08"
	32"	46"
	50"	55"
Total	8'13"	8'27"
11	1'30"	1'52"
	34"	44"
	2'05"	1'18"
	44"	56"
	1'02"	45"
Total	5'55"	5'35"
12	25"	25"
	1'24"	1'54"
	43"	38"
	17"	22"
	20"	16"
	25"	19"
Total	3'34"	3'54"
GRAND TOTAL	85'09"	85'51"

Comments on table 3

The explanations used in the explicit groups were meant to be "optimal" which simply meant that they were not going to be limited by theoretical considerations as to length or wording or by grammatical theories. When it was deemed necessary to insert an explanation, this was done, and the best wording we could produce was resorted to. Only one limitation was introduced in Ee and Es; there was an equal number of explanations coming in at the same point in the programmes. They were not translations of each other and they were not always of equal length, but there is exactly the same number of explanations. The Swedish explanations are often somewhat longer than the English ones since comparisons with Swedish were added to the comments on English usage, but this often does not show up in recording time since in giving the Swedish explanation we could often speak faster.

As can be seen from table 3 the individual explanations varied up to almost 50 % even though normally they are fairly close to each other in length. The difference in total time per lesson varied up to more than 15 % (in lesson 2), but as the table shows they add up over the total project period of 12 lessons to within 40 seconds of each other. This is partly due to pure chance, since the amount of explanations included was not pre-determined.

The number of explanations per lesson varied a lot: in lesson 8 there were only two explanations, in lesson 2 there were eight. In lesson 12 there were six explanations, all of them short: in this lesson no new stuff was introduced, and these explanations are all of the "reminder" kind.

1968, p. 43, italics ours). The Authorized Swedish Curriculum (Supplement Engelska, p. 14) also recommends that generalizations - if they are to be given or formulated at all - should come in at the end as a confirmation. This might be a point worth investigating as Smith-Baranyi (p. 113) point out, but it was not part of the present project, and we put in explanations at what was felt to be the best possible points.

The same structure was explained or commented on more than once, of course. Normally the first time was in the form of a short *eye-opener*, e.g. in lesson 10: "Now listeners, before you answer the questions I will tell you what we learn from these examples. After 'good at' we have the ing-form of the verb. So it's not enough to say 'sing' or 'swim' after 'good at'. We must say 'good at singing', 'good at swimming'." Then follows, sometimes after another short reminder, *the main explanation*, which often takes the form of a discussion, a dialogue between the voices on the tape, and with the pupils participating orally and by writing down certain phrases. Then, in a following lesson, there is a

reminder, as in lesson 11: "So, listeners, here we are going to practise sentences where we say 'afraid of'. What form of the verb must we have after 'afraid of'? //// (Pause for the pupils to think and answer) - We must have the ing-form. - Yes, that's right. Listen, please. 'He is afraid of taking the medicine' And why do we have the ing-form? //// - Well, it's because of the little word 'of'." (etc)

"Ee 7"

In lesson 7, the implicit version of which was described in detail above, explanations in the explicit versions came in at the following places. The first very short comment came in just before the pupils listened to page 2; it took 25 seconds and it pointed out that "in this little exercise we practise 'anything' in sentences with 'not'".

The next one came in just before they started writing on page 4 and it pointed out in the form of a dialogue between the voices on the tape that 'any, anything' are used in negative sentences and questions and 'some, something' in "other sentences". It took 39 seconds.

The third one, which took no less than three minutes, replaced the introduction to page 5. Instead of a mechanical but systematic discussion of all the pictures and the two things that they all expressed, a commented version, concentrating on the first two pictures and then going over the others very rapidly, was given.

The fourth and last theoretical comment in this lesson was in the form of a short reminder before the pupils started writing on page 6. It took 40 seconds. (Times given here refer to Ee; Es differs by twenty seconds only.)

The total running time of the explicit lessons (lesson 7) was about the same as that for Im (see table 1 above).

THE GUME 4 PROJECT
A DESCRIPTION OF THE LAY-OUT

Objectives.

Although the research strategy was modified on essential points as a consequence of the GUME 1-3 results (see above, p. 5), the main objectives remained almost the same:

1. to investigate what effects theoretical explanations in juxtaposition to pure structure drills may have on learning as compared to drills without explanations
2. to compare learning effects when
   a) explanations are given in the target language (English)
   b) explanations are offered in the source language (Swedish) and comparisons made with it
3. further production of various sorts of achievement tests in English
4. continued production of instructional materials.

In the main the present report will deal with points 1 and 2.

Experimental Procedure.

Schedule. The experiment was carried out in 27 classes in March, April, and May, 1970, according to the following time-table:

March 10-20: IQ testing.
31: All lesson material distributed to the schools.

April 1 + 2: Pre-test given one hour each morning.
2: Introductory lesson in Swedish given by tape-recorder in all classes, explaining experimental aims and procedure and drill techniques etc (inskolningslektion).
3: first lesson run.
6-24: (three weeks): lessons 2-10 (three each week).
27-29: lessons 11 and 12.

May 4-6: Post-test, and Attitude tests.
11-15: Standardized test.
19-22: PACT. Project ends.

May 26 and 27: Conferences with the teachers (half of them each day) and all data collected.

June 1-26: Data processed by computer (additional computer processing in August - October).

Techniqual Arrangements. All classes had a tape-recorder and a separate loudspeaker on the wall to give better sound than the built-in loudspeaker could produce. In all lessons except the last one an overhead projector was used. All classes had 12 tapes with the lessons, one with the introductory lesson, two with the achievement test; as a matter of fact the first one of these was in two different versions for the pre- and the post-tests, to give a better introduction to the two tests respectively. Apart from these 15 tapes the teacher had a large box full of pupils' lesson materials that was handed out before each lesson and afterwards collected again. The pupils were not allowed to keep any material and were not supposed to do any homework. These boxes were collected from the schools after the project.

Teaching Methods.

The experimental treatments (independent variables) used in the study (and described in the two previous chapters) are nominally the same as those used earlier, namely the Implicit and the two Explicit methods, abbreviated

Im
Ee
Es

However, since there are obvious discrepancies between these methods and those used in GUME 1-3, and since interpretation of the results is dependent on a clear picture of "what happened in the classrooms", we have given this rather detailed description of the Im, Ee and Es strategies, thereby also relating them to other strategies that have been used recently in other research projects.

The Experimental Population.

Number of school classes. At the outset it was decided that a fairly large number of school classes be used in the study. Carroll (1969)

states that it has become a sort of unwritten rule of thumb in educational and psychological research that there should be a minimum of about 20 observations within a group in order for the experiment to have sufficient power to reject the null hypothesis in a reliable way. When Carroll criticizes the Pennsylvania project (ibid., p. 216) for insufficient number of school classes in some of the comparisons, it should be remembered that *the school class mean* was the unit of analysis. If it should be maintained that the school class mean were the only acceptable unit of analysis in studies of that kind, the implication would be that no comparative field studies would be worthwhile unless at least 20 school classes (groups) were exposed to each treatment. In the case of GUME 4 this would have meant 60 school classes, an unwieldy number considering the administrative work involved and the resources in personnel and money available to the project.

The use of individual scores as the unit of analysis when the intact school class is the sampling unit is disputable since error (associated with unknown school class characteristics) is introduced. However, experiments are always a compromise between the ideal and the manageable, and the relatively large number of classes used in GUME 4 (3 x 9 = 27) is assumed to have counterbalanced this particular source of error to some extent. One reason for deciding on exactly 27 classes was the fact that an investigation of interaction between teaching method and student ability (three levels, see page 57) was planned; in the case of GUME 4 this would give a design consisting of 9 cells. With 27 school classes there is a good probability that each cell will contain a minimum of 50 students. The prejudice of some researchers, including ourselves, is that if a difference between two treatments is not clearly apparent when each treatment is applied to fifty cases, then the phenomenon is one of small consequence (cf Travers, 1960).

Selection of school classes. In November, 1969, a request for participation in the study was sent to a number of school districts. The headmasters were asked to distribute the request to all teachers in grade 6. The teachers were required to fill in a questionnaire about prevalent conditions in the class (textbook used, instructional aids available, discipline, the teacher's experience in that particular class, the number of pupils, etc.). A surplus of teachers willing to participate was obtained this way, and experimental classes were chosen among those using one particular textbook (Ashton-Olsson, "Hands up"), which

was our first prerequisite for participation, and showing the greatest conformity in a number of characteristics (according to the teachers' responses to the questionnaire). A list of participating school classes will be found in Appendix E. All the classes are from Gothenburg though with a large overrepresentation of classes from the western and northern parts of the city, whereas GUME 5 (see p. 6) used classes and schools in other parts of the city.

Assignment to treatments. The 27 classes were randomly assigned to teaching methods. However, one restriction was applied to this procedure: no two classes from the same school were allowed to get the same treatment. Incidentally, the randomization procedure was undertaken on March 12th, 1970, shortly before the beginning of the project and after all materials were written and the teachers informed about the project.

Drop-out rate. In the participating school classes there were altogether *685 pupils*. However, 65 of them missed either the Pre-test or the Post-test and were eliminated from the data processing, thus leaving *620 pupils*. Of these, 43 pupils were absent from more than two lessons during the experiment and were cancelled from the computations, which leaves *577 pupils* for the experiment, and it is always this group that we refer to later on. Concerning the two types of drop-outs, a word of comment may be appropriate:

a) For those pupils who were absent on the Pre- or Post-test ($N = 65$), no data cards were punched. Although information on a number of variables was available about these pupils, they could not be used in the main investigation (treatment comparisons) and were therefore dropped.

b) The pupils who were excluded because of too high a rate of absence ($N = 43$) will hereafter be referred to as *the drop-outs*. They will be compared on a number of background variables with the experimental population to find out whether they deviate in any systematic way from the main population.

EVALUATION INSTRUMENTS

The Achievement Test.

In any experiment of the kind that GUME represents the results are dependent on the test used to measure progress. If the test is not sensitive enough to measure differences that do exist, or if it is biassed one way or another, so that one of the strategies under investigation is favoured, all conclusions are invalidated.

Progress, i.e. the difference between what the pupils knew at the beginning of the project and what they knew at the end of it, was measured by a 160-item, 80-minute achievement test, specially made for the project. Its validity and reliability, which are important, will be discussed below (p. 48).

Since the Achievement test is a written test, it may be argued that an important aspect of language mastery, namely the spoken language, was unduly neglected. A word about that may be in order.

It should be stressed first that we had never planned to cover the whole field of language learning; we were only interested in the pupils' active mastery of certain grammatical structures (in speaking, writing, and reading).

We did use an audio comprehension test, PACT, described below, and in the Standardized test there is a listening comprehension test and a pronunciation test (although "silent", see below). Moreover: in marking our tests spelling mistakes were overlooked if they seemed to indicate a correct spoken form, e.g. 'like's' and 'das' for 'likes' and 'does'. We felt that by doing this we did in fact, to a certain extent, measure oral performance: if the pupils knew how to say the phrases, they should be able to write them well enough for the marker to see this and to give full points.

Developing more sophisticated oral tests, which is a most important task for many reasons, was not within the scope of the present project, nor was it financially possible. This is one of the prime tasks of the continued GUME activity (see fig. 1, p. 7). There is, in all likelihood, so high a correlation between that kind of test and the one used in the present study, especially when our generous kind of marking is used, that the introduction of it would have made little or no difference

for our purposes. As a long-time influence on teaching, however, it will be of paramount importance.

Some parts of the Achievement test in GUME 4 were used in the previous studies in 1968-69, and after careful analysis of items they were re-written. Three consecutive versions of the test were tried out in a number of classes, item-analysed and re-written before the test got its final shape.

The twelve parts of the test (see Appendix A) will here be commented on briefly.

Test A. Ten items testing the pupils' active correct use of the s-form of verbs in the third person singular present tense. Spelling mistakes were overlooked as described above.

Test B. 15 items testing ability to form correct questions with main verbs in the present and past tenses. Minor spelling mistakes overlooked.

Test C. 45 items. This is in fact a multiple-choice test, but the alternatives are all given at the beginning of each little part of the test. This arrangement, which makes it more like a completion test, was adopted partly because of the wide-spread critical attitude among language teachers against multiple-choice tests.

This is a multiple-purpose test also; we try to test primarily the ing-form after prepositions and the correct use of the present and past continuous as opposed to the simple present and past. There are also some examples of the s-form as a result of this. It turned out in early stages of the experiment that it is difficult to construct a good test of, for example, only the ing-form after prepositions since this tends to give the pupil either all correct or all wrong. Therefore this mixed test. We have later marked it for the various aspects that it sets out to test, and thus there are separate results for preposition followed by an ing-form and also for the s-form of verbs. (This will be discussed further under the heading "Critical Items", p. 89 below.)

Test D. 20 items testing the correct position of adverbs of time in connexion with main and auxiliary verbs. The problem for a Swedish child arises from that fact that in Swedish such adverbs are placed after all verbs, including auxiliaries, in main clauses. In a preliminary version of the test we used squares before and after the verbs, where the pupils were asked to put a cross. This made the test a two-choice one with too little spread in scores.

Test E. 15 items testing ability to form correct questions. This test is an exact parallel to test B except that the stimuli are different: here the pupils are asked to make a transformation whereas in test B they were told to ask the question. Marking followed the same principles, and the results on the two tests should be compared.

Test F. This is a 40-item 6-option multiple-choice test of the some-any dichotomy also testing the semantic difference some-somebody-something. Care has been taken to include only unambiguous examples and it was tried out on a number of people speaking English as their native language. There are also a total of 15 "critical items", i.e. questions and negative sentences with 'some' and statements with 'any', like 'Don't forget to write some letters!', 'It could happen to anybody' These have been investigated separately (see p. 91 below).

Test G. 15 items testing ability to form correct negative sentences with main verbs. This test is no doubt valid in relation to the kind of teaching that was offered in the project and also the teaching the pupils usually meet in the classroom, but it may have been technically somewhat complicated for some pupils, and there is a risk that some did not understand what they were supposed to do. Most of them did, however, and the low scores are due to the fact that they formed incorrect questions without any auxiliary 'do' or with the wrong form of it.

Marking and test administration. The marking of all tests was done by student teachers from the Gothenburg School of Education, all with some experience of teaching English at the level under investigation. Parts C, D, and F were marked according to a right-wrong pattern with no other risk of mistake than ordinary human error. The remaining parts were marked as right or wrong (no half-points were given) according to careful instructions which have been discribed above, and all uncertainties were discussed with the project staff. Inter-scorer reliability coefficients have not been calculated (each test has only been marked once), but careful scrutiny of a number of tests have not revealed any mistakes except a few simple oversights.

The test was administered in two different class periods given on two days following each other, usually some time between 9 and 11. All instructions were recorded on tape which ran through the whole testing period and thus was responsible also for timing the test. The teachers were not allowed to give extra help or instructions. All instructions were in Swedish.

Validity and reliability. The validity of the achievement test, i.e. the extent to which it measured what we wanted it to measure, has been established through correlations between the test and grades given by the teachers and also results on the Standardized test, which is an established norm for knowledge of English at this level. These figures are given in table 33 on page 98 below; the correlations Pre-test - grades in English are .70, and Pre-test - Standardized test total .83. These figures show that the test has high validity and can well be used for its purpose as far as the contents is concerned.

A subjective estimate of content validity may be in place since the goal of the project is not primarily knowledge of English in general but knowledge of certain grammatical structures in particular. This can be done by a comparison with the goals as set down in the Curriculum for Swedish Schools (Läroplan för grundskolan) and with the contents of the textbooks used at this level, especially the one used by all experiment classes, "Hands up" by Ashton-Olsson. Such a comparison shows that all structures taught in the project take up a central position in the course for our pupils; some of them make up the very backbone of the first three years of any course in English as a foreign language, the others, like preposition followed by an ing-form of the verb, the position of adverbs and the some-any dichotomy, are all important ingredients of the intermediate level course that the pupils reach in the 7th form, just after the summer holiday following the experiment.

The validity of the test used must thus be considered quite satisfactory.

The reliability of the test, i.e. the accuracy and constancy with which it measures, has been calculated by the Kuder-Richardson formula 21

$$\frac{n}{n-1} \cdot \left(1 - \frac{\bar{x}\,(n - \bar{x})}{n \cdot s^2}\right)$$

where n is the number of items in the test (see Thorndike-Hagen, p. 185). The reliability coefficients obtained for the Pre-test were those given in table 4.

Since a reliability of about .50 is enough for group comparisons, which is what we are concerned with here, the figures for the test and its parts are very satisfactory. The total is good enough even for diagnostic and prognostic purposes with individuals, for which figures around .90 and .95 respectively are required.

Table 4 : Reliability Coefficients for the Achievement Test (Pre-test).

Part	
A	.41
B	.78
C	.75
D	.83
E	.78
F	.79
G	.81
Total	.94

The Pupil Attitude Test.

The pupils were given an attitude test at the end of the project. This questionnaire is given in Appendix B, but it will be discussed briefly here.

The questionnaire consisted of two parts. In the first part the pupils were asked to state their interest in all subjects they have on their time-table in the 6th form in one out of four categories: almost always fun, more fun than boring, more boring than fun, almost always boring. Figures for English are discussed in this report on page 108, all other subjects in Appendix C.

The second and main part of the questionnaire consisted of 12 questions on various aspects of the project. The first two questions were open: What was good, what was not very good with the project was that ... . These questions were put at the beginning to get spontaneous reactions from the pupils. Then there were directed questions with four or five optional answers asking the pupils how much they felt that they had learnt, how they had liked the lessons, whether they felt that they had understood what they had been doing, whether they felt the explanations had helped them, or - in the Implicit groups - whether they had missed explanations, and finally there were four questions on the four-phase drills which were probably new to most pupils.

The teachers had been asked not to discuss the project with their classes until after they had filled out this questionnaire so that teacher attitudes would not influence the pupils. How much they had been discussing between themselves we cannot know, of course.

The Teacher Attitude Test.

At the end of the project all teachers were given a questionnaire in which they were asked to give their attitudes to a number of points in the project. It is given in Appendix D, but a short description of the questionnaire will be given here.

In a first part of it the teachers were asked half a dozen questions on how they usually teach English themselves, which method they use (as compared to those used in the project), how they treat grammatical difficulties, how much they speak English, and whether they use structure drills.

In the second part of the questionnaire, which makes up 2.5 of its 3 pages, they were asked to comment on the project, what they liked and what they did not like. They were asked to comment on oral drills, written exercises, reading texts, explanations, tempo, technical quality and problems, and they were also asked to estimate the learning results in the pupils (we wanted to compare the teachers' subjective estimates with the objective findings afterwards; this also turned out to be quite interesting). They also commented on the test and - some of them - on individual lessons.

The questionnaire was mostly in the form of open questions which makes it somewhat difficult to tabulate, but this was done on purpose so as not to direct their answers one way or another more than necessary.

The Standardized Test.

All Swedisn students in the sixth form (age about 13) are given standardized tests in Swedish, English and Mathematics, prepared by the National Board of Education. The English test, new norms for which were worked out in 1969, has been used for many years and is somewhat out of step with recent developments in language instruction. The test is normally given between April 14 and May 9, but in all project classes it was given between May 10 and 16 so as not to interfere with the project. The test consists of four parts:

EL (Engelsk läsning), English Reading. A reading comprehension test consisting of nine short texts varying between 4 and 14 lines in length followed by two, three or four questions, with a total of 24 questions. This is then a multiple-choice test with four options to each question. It takes 35 minutes effective working time.

EM (Engelska meningar), English Sentences. This is a fill-in test consisting of 26 sentences meant to test "the pupils' general linguistic feeling and their knowledge of simple grammatical phenomena". In reality it is a grammar test only testing knowledge of simple accidence ('formlära'). The 'basic form' (infinitive of verbs and singular of nouns) is given and the pupil is asked to fill in the appropriate inflected form. There are no less than eight irregular verbs in the past tense but no example of the do-construction in either a question or a negative sentence. Many modern textbooks, including the one used by all experimental classes, take up irregular verbs for systematic study so late in the sixth form that many classes have not had time to deal with this problem before they do the test. They have all worked with the do-construction for almost two years. This is a drawback of the test which negatively influenced the results in all experimental classes since none of them had had time to deal with irregular verbs. This test takes 8 minutes.

EA (Engelsk avlyssning), English Listening. This is a listening comprehension test. There are 24 items. The test is recorded on tape. The pupils hear a sentence or two followed by a question or sometimes just a question and on his answer sheet he has five options to choose between. One of the examples before the test starts: 'What do I put in my tea?' Options: butter - cheese - pepper - sugar - ice. The test is meant to test the pupils' ability to understand spoken English. Many of the items are as much or even more tests of vocabulary since what is spoken is very easy to understand but the options are difficult to choose between unless you know the words well. This contamination with the written language is common to oral tests. (PACT, the Pictorial Auditory Comprehension Test used in the project (see below) is an attempt to get away from this problem.) The test takes 12 minutes, and it is given together with EU in a second testing period.

EU (Engelskt uttal), English Pronunciation. This is a "silent" 24-item pronunciation test. The pupil sees a key word with one "sound", a vowel or a consonant represented by one or two letters, italicized and he is asked to choose one out of five options that contains the same sound (but the words do not rhyme since the sound can occur in different positions in the two words). An example: 'early: short - green - girls - great - ready'. It is pointed out in the Instructions booklet that this test measures "primarily the pupils' control of the individual sounds of the language. The test has proved to correlate highly with the pupils' general pronunciation of English". This test takes 14 minutes.

The four tests are given in three different class periods, normally on three different days. The times given above do not include instructions, and three full periods are normally used for this test battery. The pupils can score a total of 98 points which are then transformed by the teacher according to norms into a 5-point grade scale. The tests are not constructed to help the teacher grade individual pupils but to give him an idea of the general standard of his class.

The reliability coefficients as calculated for the pupils within the present project are (K-R formula 21): EI (.74), EM (.84), EA (.82), EU (.78), Total (.93).

PACT.

The original test, called Pictorial Auditory Comprehension Test, was developed by John B. Carroll and one of his assistants, Wai-Ching Ho. It is a listening comprehension test intended to measure foreigners' comprehension of spoken English. In the earlier GUME experiments mimeographed copies of the original version were used with kind permission of Dr. Carroll. In the present study, however, an entirely new version was worked out, although with the original testing technique preserved. Thus the pupils listened to a taped conversation or description of an object or event, etc., and then marked which of four alternatives (in the form of pictures) corresponded to what was said on the tape. The test consists of 55 items and takes 30 minutes to administer. The reliability (K-R 21) of the test is .85.

As was mentioned earlier in the report, test development is one of the objectives of the project. Although auditory tests have been available in the Swedish schools, none has been uncontaminated as far as reading ability is concerned (the options on the answer sheet have mostly consisted of w r i t t e n alternatives). PACT seemed to be promising in this respect and was therefore investigated in the project. The test will be further commented on in the Results section. On the next page the test technique is illustrated by an example.

The Intelligence Test.

In the present study the same test as was used in GUME 1-3 was administered, namely the verbal, inductive and spatial factor tests of the so-called DBA-test (DBA = Differentiell BegåvningsAnalys,

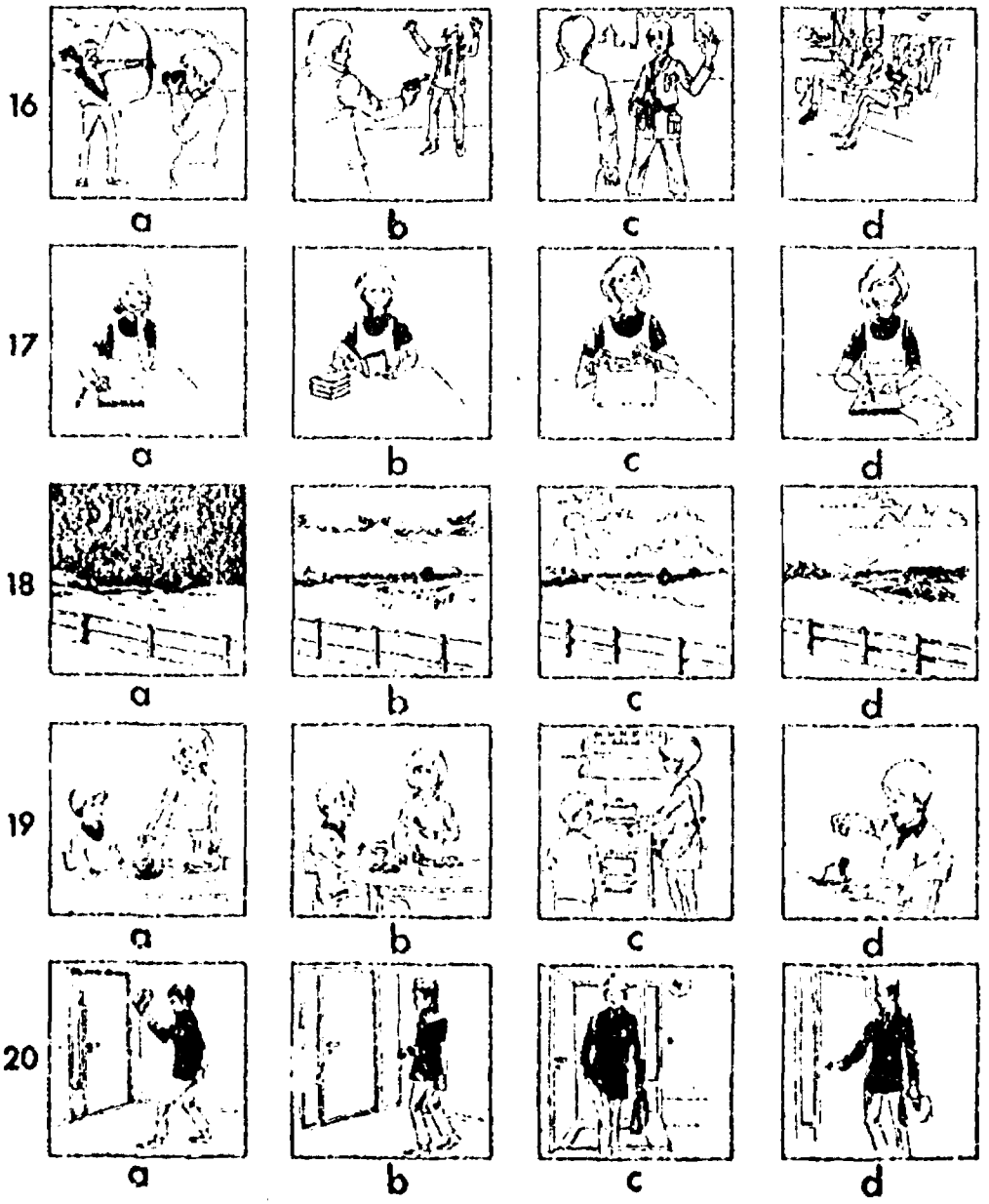


Fig. 5: Example of Page in PACT.

PACT.

In the fig. above items nos. 16-20 of the test are given. As a typical example the auditory stimulus of item no. 16 is given (the following is said on the tape):

"You must stand very still. It's so dark here that I have to take a long time, and if you move the result will probably be a bit shaky".

The pupils mark their answers on a separate sheet.
It's a which is correct, of course.

i.e. differential intelligence analysis) constructed by Professor Kjell Härnqvist of the University of Gothenburg. The three subtests, taken together, are considered to be a reliable measure of general ability or scholastic aptitude (see further Härnqvist, Manual till DBA). The sum of the pupils' stanine scores were transformed to T-scores with a theoretical mean of 50 and a standard deviation of 10.

The tests were given approximately two weeks before the experiment proper started; they were administered on the same occasion and in the following order: Verbal (10 min.), Inductive (15 min.), Spatial (12 min.).

Other Measures.

Social class. Information about the parents' occupation was collected at the headmasters' offices. The intention was partly to check the social background of the different treatment groups, partly to investigate the correlation between this variable and others used in the study. The criterion for assigning a pupil to a particular social class was a hierarchical description of professions and occupations from 1958 (1958 års valstatistik), which is to some extent arbitrary and even inconsistent, but it is the only source available at the moment. Social class 1 corresponds roughly to English "upper middle class", and class 3 to "working class"; the much disputed division is based mainly on income only. A zero was used as the code for cases where the mother (without any mention of profession) was given as the guardian in order to make further analyses of this group possible.

Grades. Grades in English, Swedish and Mathematics were collected. The grades had been given at the end of the preceding term, i.e. the autumn term, when the pupils were in their first term of the sixth school year. It should be noticed that the grades had not been corrected or adjusted according to any standardized achievement test, simply because no such had yet been given. Thus the grades reflect a relatively great subjectivity. They are expressed on a 5-point scale (theoretical mean 3 and standard deviation 1). The three grades were added together whereby a scale with a standard deviation of 3 was obtained. Since the intention was to give IQ and Grades equal weight in the statistical analyses, the grade score was multiplied by 3. The Grade scale, thus obtained, had a mean of 27 and a standard deviation of 9 (equality of weight is dependent on the standard deviation, not on the mean).

DATA TREATMENT

The Statistical Program.

All data were processed at Göteborgs Datacentral för Forskning och Högre Utbildning by computer IBM 360/65. Statistical programs included in the ISR (Institute for Social Research, University of Michigan) and BMD (Bio-Medical Computer Programs, UCLA) series were used. The following measures or analyses were obtained:

a) *Means, standard deviations and frequency distributions* for all variables. These data were obtained for the total population, for boys and girls separately, for the three treatments (Im/Ee/Es) separately and for each participating school class.

b) *Correlations* between all variables for the whole group.

c) *Analyses of variance (one-way)* of a number of independent variables in order to investigate comparability between treatment groups (three cells).

d) *Analyses of variance (two-way)* with the experimental population divided into three levels of intellectual ability (nine cells).

e) *Analyses of covariance* with different covariates and dependent variables.

The purposes of the various analyses will be given below. Any pupil not attending ten or more lessons was eliminated from the data processing. In a field study of the present kind it is necessary to accept a certain amount of absence. The decision to draw the line at 2 lessons is a matter of subjective judgment though it is probably the most realistic value considering availability of subjects. Those pupils who did not take the Pre-test and the Post-test were also eliminated from all computations, even if they had taken part in the whole lesson series. Within the experimental population the N's vary somewhat from variable to variable due to stray absences.

Experimental Design.

The design used corresponds to Campbell and Stanley's "design 10", the Non-equivalent Control Group Design (Gage, 1963, p. 217). For administrative reasons intact school classes had to be used in the

experiment. It has thus not been possible to assign pupils randomly to teaching strategies (treatments). In the absence of experimental control of background (concomitant) variables, statistical control by analysis of covariance has been resorted to when investigating the main effects.

The unit of analysis used in the study is the individual score. Since it might be argued (see p. 43) that the school class mean would be the proper unit of analysis, an investigation of main effects has also been made with the school class means as the units of analysis. Of course, with such a limited number of school classes as are used in the present study, the loss of degrees of freedom is great when the analysis moves from the individual to the school class level.

Computation of Main Effects.

The main purpose of the experiment is to investigate which of the three teaching methods produces the best learning result. The measure of progress that was used throughout the computer analyses was the difference in raw scores between the Post-test and the Pre-test. In addition, two other measures of progress were used though in those cases the computations were made by hand. The particular measures will be presented below.

When the three teaching strategies were compared with respect to Progress, the following covariates were used in the analyses of covariance: IQ, the Pre-test, the Standardized English Test, and PACT. The four measures were used separately in four different analyses; in a fifth analysis they were weighted together to a composite measure. Treatment effects were also compared with respect to Post-test scores; in this case the Pre-test served as the covariate. The analyses of covariance mentioned thus far may be summarized thus:

Table 5 : Analyses of Covariance Performed.

Dependent variable	Covariate
Progress	IQ
-"-	Standardized English Test
-"-	PACT
-"-	Pre-test
-"-	The above four weighted together
Post-test	Pre-test

Computation of Interaction Effects.

Since it can be hypothesized that one particular teaching method facilitates learning for one particular ability level more than for another, the interaction between teaching method and ability level was investigated. Analyses of variance, two-way classification, were performed with Progress and the Post-test as dependent variables. The experimental population was divided into three equal parts according to scores on the IQ test. The IQ scores (ranges) for the lower, middle and upper third respectively turned out to be: 29-49, 50-58, 59-77. The data were organized in a 3 x 3 table in each analysis, thus:

	Im	Ee	Es
Upper			
Middle			
Lower			

Retention.

According to the original research plan the Achievement test should be administered a third time, when the pupils were just starting grade 7, in order to measure retention or, rather, differential retention between the three methods. (In GUME 1-3 the retention tests were given one month after the experiment). However, for the results to be interpretable it would have been necessary to control the teachers for an unduly long period of time, preventing them from teaching the structures dealt with in the project. Since it was considered unrealistic to control the teaching process thus, the retention test was dropped.

Various Measures of Progress.

As has been mentioned earlier, the pupils' progress during the experiment was measured by the difference in *raw scores* between the Post-test and the Pre-test.

However, it may be argued that a measure of progress must somehow take account of the pupils' standing on the Pre-test. If, for instance, a pupil scores very high on the Pre-test, there is not so much room for progress because of ceiling effects. The following index takes care

of this, giving more weight to progress scores "at the upper end of the scale":

$$\frac{\text{Actual improvement} \times 100}{\text{Possible improvement}} = \%$$

An example: Pupil A has 100 points on the Pre-test and 120 on the Post-test, pupil B has 80 on the Pre-test and 100 on the Post-test. The raw progress of both these pupils is thus 20 points and according to this measure they have made the same progress. The Achievement test has a maximum score of 160. Possible improvements for the two subjects are 60 and 80 points respectively, and their scores as computed by the above formula then become 33 (%) and 25 (%) respectively; thus according to this measure pupil A has made greater progress.

On the other hand it may be argued that increments among inferior pupils are of greater consequence than equally great improvements (in raw scores) among superior pupils. However dubious this way of reasoning may be, the following index of progress gives higher credit to improvements "at the lower end of the scale":

$$\frac{(\text{Post-test} - \text{Pre-test}) \times 100}{\text{Pre-test}} = \%$$

Both these measures have been calculated with school class means as the unit of analysis.

STATISTICAL DESCRIPTION OF THE EXPERIMENTAL POPULATION

Attendance.

The pupils who were absent more than two lessons during the experiment were dropped from the computations. In the table below the experimental population is described with respect to attendance during the series of lessons.

Table 6 : Attendance of the experimental population during the series of lessons.

	Number of lessons attended			
	12	11	10	
Boys	185	66	24	275
Girls	212	62	28	302
Total	397	128	52	577

For the purposes of the experiment, the pupils who were absent one or two lessons (N = 180) were considered comparable to those who had 100 % attendance. As a partial check on this proposition, absence was included as a variable in the calculations of correlations. As it appeared, absence (defined as absence during 1 or 2 lessons) did not correlate with any other variable.

Assignment to Treatments.

Since the school class was the sampling unit and since the boys/girls proportion varied from class to class, the distribution of the sexes on treatments was a matter of chance. The actual distribution is presented in the following table:

Table 7 : Distribution of Pupils on Teaching Methods.

	Im	Ee	Es	Total
Boys	98	90	87	275
Girls	83	105	114	302
Total	181	195	201	577

As is apparent from the table, the boys/girls ratio within the Implicit group deviates from that of the others. However, a X^2- test shows that the deviation is not statistically significant ($X^2 = 4.99$, $df = 2, p > .05$).

Nor does the observed number of pupils (disregarding sex) per method deviate significantly from what is theoretically desirable. ($X^2 = 1.15$, $df = 2, p > .50$).

Social Class.

The distribution of the experimental population on social classes is given in table 8.

Table 8 : Distribution according to Social Class, Absolute Figures (N = 577).

	No inform.	0	1	2	3	
boys	9	25	21	96	124	275
Girls	20	26	20	108	128	302
Total	29	51	41	204	252	577

The 0 stands for cases where the mother is responsible for the care of the child. Although this group is probably very heterogeneous with respect to social class (ordinarily no information about the mother's

occupation was available at the headmaster's office; when it was, the pupil was assigned to the corresponding social class, however) it was considered of interest to investigate this particular group with respect to certain variables.

In the following table the pupils mentioned above and those for whom no information was obtained, have been disregarded. The remainder, i.e. those in social classes 1, 2, and 3, have been transformed into percentages.

Table 9 : Distribution according to Social Class. Percentages (N = 497).

	1	2	3	Total
Boys	4.2	19.3	25.0	48.5
Girls	4.0	21.8	25.7	51.5
Total	8.2	41.1	50.7	100.0

The experimental group is very close to the "norm" with respect to social class distribution. According to official statistics for Gothenburg (Andrakammarvalet i Göteborg 1968, U 1969:2, pp. 63-69) the overall figures for social groups in Gothenburg are:

1: 8.2 % 2: 38.4 % 3: 53.4 %

The deviation from this norm was tested for significance. The X^2-value obtained was 1.54 with 2 df, thus being far from significant.

Course Choice.

In February all pupils in grade 6 in Sweden had to choose which of the two courses (sk and ak) they wanted to take in grades 7 through 9. The choices that our experimental pupils made are presented in the following table:

Table 10: Distribution according to Course Choice in English for Grade 7.

	No inform.	ak	sk	Total
Boys	2	80	193	275
Girls	1	62	239	302
Total	3	142	432	577

Discounting the pupils for whom no information was available we find that 24.7 % of the pupils chose the easier course (ak) whereas 75.3 % preferred the more advanced one. These figures deviate somewhat from those of grade 6 at large, which proved to be 29.5 % and 70.5 % for ak and sk respectively (Information from the Gothenburg Board of Education). It should be noticed, however, that the figures in GUME 4 were based not on official statistics from the headmasters' offices but on the pupils' reports some time after the formal choice was made. It could be that some pupils had forgotten the actual choice or that their memory was selective in this respect (assuming that it might give more status to take the advanced course). All in all, the information is probably somewhat unreliable and the ak/sk variable should be treated with some caution.

Representativity of the Experimental Group in Certain Variables.

As a further check on the representativity of the experimental population its standing on a number of well-defined measures were gathered. They will be given below. If the results of a study such as the present one are to be generalizable, it is necessary that the population on which the treatments were applied is not atypical.

IQ. In the table below results on the three parts of the DBA-test are given. Values are given for boys and girls separately.

Table 11 : Means and Standard Deviations on the DBA-test (the parts in stanine points and the total in T-scores).

	Boys			Girls				
	N	$\bar{x}$	s	N	$\bar{x}$	s	t	sign.
Verbal IQ	269	5.17	1.78	295	5.42	1.79	- 1.66	
Inductive IQ	269	5.85	2.04	295	5.74	1.83	.67	
Spatial IQ	269	5.75	1.84	295	5.39	2.07	2.18	x
Total	269	53.91	9.58	295	53.43	9.70	.59	

The DBA-test was originally standardized in 1958. However, it has been noticed in recent years that the norms developed then have become outdated. A new standardization was therefore undertaken in 1967/68 (Härnqvist, 1969, a and b) and new norms were established. As it appeared a certain increase in raw scores was found with respect to the verbal, inductive and spatial factors, i.e. the same variables as were used in the present study. (In the case of the numerical and perceptual variables, which were not used by us, a decrease was noticed). Furthermore sex differences appeared to diminish from 1958 to 1967/68 with respect to all variables. In the revised test manual, the increase in raw scores has been taken into account. The correction technique as well as new and old norms are given in Härnqvist (1969, a).

The figures in table 11 are thus inordinately high in comparison with the original norms, giving the impression that our sample is extremely biassed. Even after adjustment for outdated norms, however, the GUME boys seem to be a select group. The girls, on the contrary, appear to be an unbiassed sample. However, it should be taken into consideration that the sample is taken in its entirety from a large city, which of course makes comparisons with the DBA norm group dubious. Härnqvist (1969, b) refers to some investigations where the samples were recruited from larger cities, among them one with pupils from Gothenburg only (Larsson & Sandgren, 1968). When we compare their values with those of GUME 4 in variables that were the same in the two projects, namely the verbal and the inductive, we find that the boys in the two studies are almost identical, whereas the girls in GUME seem to be somewhat inferior to those of the Larsson & Sandgren study. All in all, the GUME 4 group is somewhat biassed as compared to the

national norm but corresponds well to available norms for Gothenburg. As far as general intelligence is concerned, the experimental population is such as to warrant generalizations to other large city groups.

Concerning sex differences there is a tendency for girls to excel verbally and for the boys to do better on the spatial test, which is according to earlier findings (see,for instance, Anastasi, 1958). In the case of the spatial test, the difference in favour of the boys is statistically significant. As regards the total IQ measure no differences between the sexes are found.

Grades. The grades referred to in this study had been given the preceding term, i.e. when no national norm was available to the teachers (the standardized test is not given until the spring term of grade 6).

Table 12 : Grades: means and standard deviations.

	Boys			Girls				
	N	$\bar{x}$	s	N	$\bar{x}$	s	t	sign
Grades Swedish	274	2.91	.92	299	3.37	.86	- 5.90	xx
Grades English	275	2.88	1.06	301	3.28	.97	- 4.71	xx
Grades Maths	275	3.09	.98	301	3.07	.96	.24	
Grades Total	274	2.96	.98	299	3.24	.94	- 1.10	

xx = significant at the 1 % level

The results for both sexes are according to expectations. As a matter of fact the figures correspond almost exactly to those of GUME 1-3, both for boys and girls. It is a well-attested fact that girls excel in the case of grades (see for instance, Anastasi, 1958, p. 492 ff). The superiority of the girls is statistically significant in the case of Grades Swedish and Grades English, i.e. the school subjects corresponding most closely to the verbal test. It is a common teacher experience that the grade point average in Swedish schools is now - as in our group - above the theoretical mean of 3.0. It may be stated with confidence that the experimental group is normal with respect to grades.

The Standardized Test in English. According to the norm table for this particular test the theoretical mean is 56.0. The latest check on this norm was made in the spring of 1969 when the empirical value obtained was 55.8 (s = 19.5). No norms are available for the subtests, nor are there any for boys and girls separately. However, the following table gives the values for both sexes in the GUME sample.

Table 13 : The Standardized Test; means and standard deviations.

	Boys			Girls				
	N	$\bar{x}$	s	N	$\bar{x}$	s	t	sign
EL	273	11.83	4.77	296	12.33	4.84	1.24	
EM	273	11.20	6.34	296	13.66	6.06	4.76	xx
EA	273	15.27	5.91	296	16.56	4.97	2.79	xx
EU	273	11.65	5.10	296	14.17	5.03	5.94	xx
Total	273	49.95	18.65	296	56.73	18.14	4.39	xx

The value for the total group on the whole test is 53.48 (s: 18.68) which corresponds to a grade mean of 2.90. Thus with respect to proficiency in English, in so far as it is measured by the present test, the experimental group is below the norm for grade 6. The girls are significantly ahead of the boys on the total test as well as on most subtests, which is in line with expectations.

To sum up:

The representativity of the experimental group has been investigated with respect to general intelligence, grades and achievement on the standardized English test. In all these respects the girls seem to be an unbiassed sample of the population at large (girls in grade 6) whereas the boys deviate somewhat in general intelligence and on the standardized test. In the case of intelligence, the boys are slightly above the norm, in the case of the standardized test slightly below. The total group is considered sufficiently representative for results to be generalizable to pupils in grade 6.

Characteristics of the Treatment Groups.

Earlier we found (p. 60) that the experimental population was distributed evenly between the teaching methods and that the boys/ girls ratio within methods was approximately the same. It is also of interest to investigate if the three groups are comparable in the background variables used as control measures (covariates) in the forthcoming analyses. The comparisons were made by analyses of variance; the results are given in the table below.

Table 14 : Analyses of Variance (one-way) of Certain Background Variables.

	Means				Sum of squares		
Variable	Im	Ee	Es	F	between	within	df
IQ total	53.26	54.25	53.43	.564	105	52179	2/561
Grades total	27.84	28.19	27.82	.139	17	34005	2/570
Pre-test	49.24	53.14	52.28	1.792	1560	248961	2/572

In no case is a significant F-ratio obtained. Thus the three treatment groups seem to be of equal standing as far as general intelligence, grades and pre-test achievement are concerned; not even the fairly large difference on the Pre-test is significant. One tendency is found among the figures, namely for the Ee group to be slightly ahead of the others.

MAIN RESULTS

Overall Progress during the Experiment.

A necessary prerequisite for studying differences in progress between teaching methods is that the treatments have had measurable or, preferably, substantial effects on the pupils. In other words, did the pupils, irrespective of teaching method, learn anything from the twelve lessons? The figures in the following table give a first rough answer to this question.

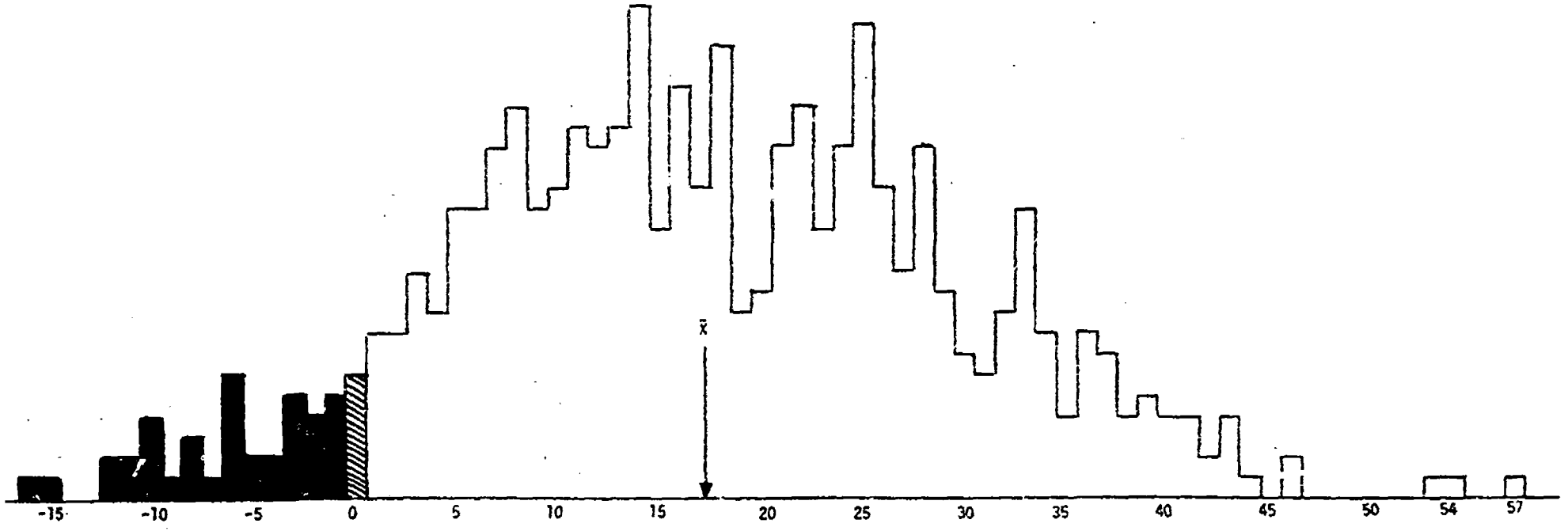
Table 15 : Pre-test, Post-test and Progress; Means and Standard Deviations.

	Total			Boys			Girls		
	N	$\bar{x}$	s	N	$\bar{x}$	s	N	$\bar{x}$	s
Pre-test	575	51.61	20.89	273	49.26	21.35	302	53.74	20.27
Post-test	576	68.67	27.16	274	64.07	27.20	302	72.84	26.48
Progress	574	17.26	12.32	272	15.21	12.15	302	19.11	12.19

The progress for the total group is substantial. There is obviously room enough for true method differences, if any, to appear. The girls are ahead of the boys with respect to pre- and post-test scores and with respect to progress. In all three cases the differences are significant at the 1 % level or very near it (t-values: 2.56, 3.91, 3.84 for the Pre-test, the Post-test and Progress respectively).

As is apparent from figure 6 , the variation in progress scores is very large. It should also be observed that the values at the negative extreme of the distribution are dubious; a negative progress, i.e. a regress, of 16 or 15 points (two pupils) is hardly a true regress but some sort of test effect, caused by failing motivation at the time of the post-test. Probably most of the regress scores, (black field) are test effects of one kind or another. On the other hand it might be argued that values at the positive extreme have been analogously caused by low motivation on the pre-test occasion. This seems less probable, however, and therefore some of the extreme regress

Fig. 6: Distribution of Individual Progress (Including Regress)
(N = 577)



scores may have been cancelled, but this was not done. Thus all progress scores (i.e. including regress scores), whatever their nature, were included in the analyses.

Progress - Main Effects.

The main objective of the present investigation is to shed light on the question: Which of the three methods, Im/Ee/Es, produces the best learning effects? This chapter contains a number of statistical analyses; before presenting them, however, we shall discuss a figure intended to visualize the outcome of the study.

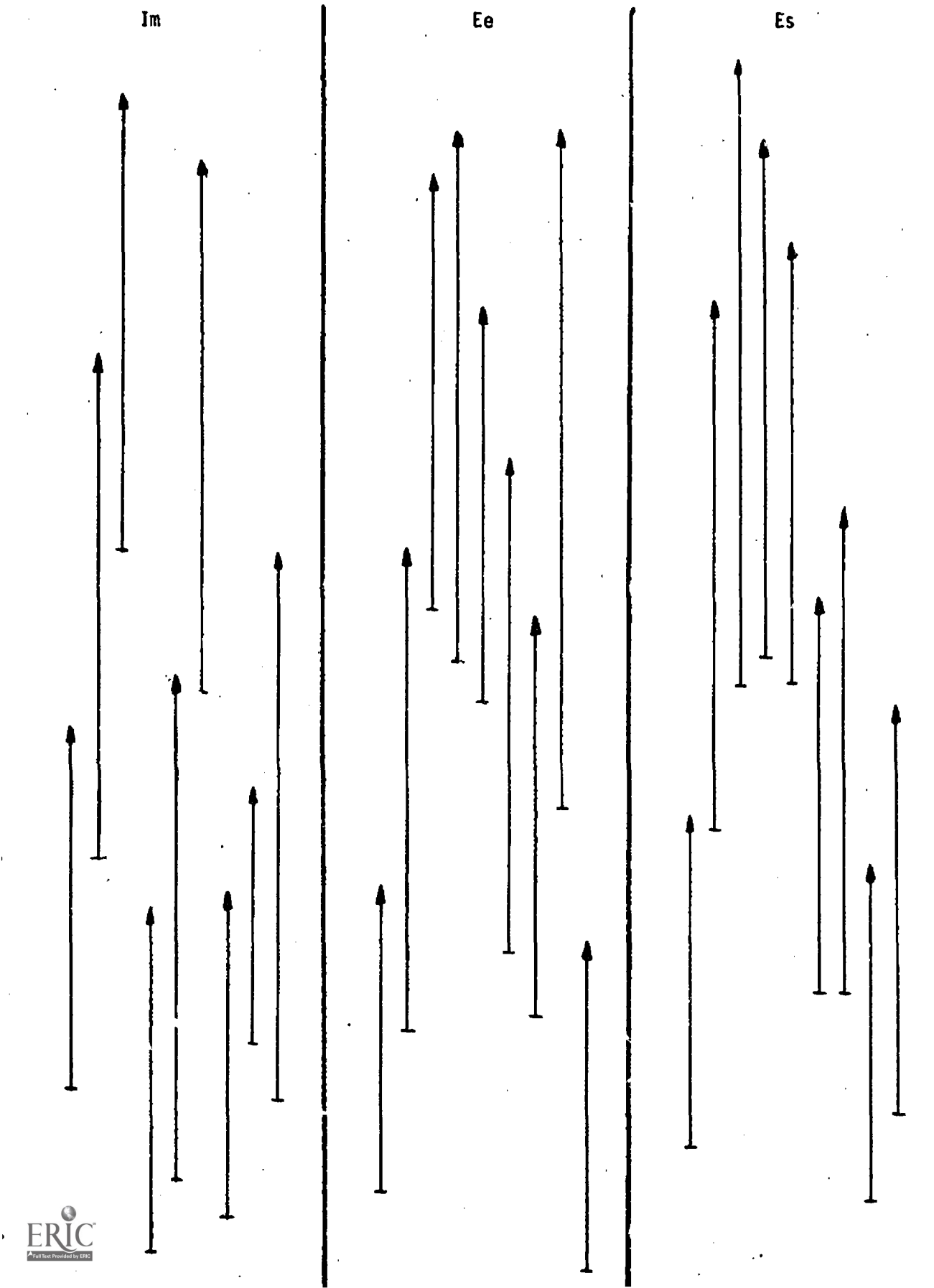
School class progress. In fig. 7 the twenty-seven school classes are indicated by arrows. The bottom end of each arrow signifies the Pre-test score, the top end gives the Post-test score and the length of the arrow is an indication of the magnitude of the progress. (Pre-test, Post-test and Progress means and s's for each school class are given in Appendix F.) The arrows are arranged in three groups, one for each teaching method.

Although the school class data give a levelled-out picture of the results, the overall impression is nevertheless one of great variation within rather than between methods. It is an interesting finding *per se* that school classes vary so strongly; as a matter of fact, the Pre-test scores of many classes surpass the Post-test scores of others. Progress (= length of arrow) is also found to vary a great deal between classes. We find the shorter arrows towards the bottom of the figure and the longer arrows towards the top, which is equal to stating that there is a correlation between school class pre-test scores and progress; the better the class at the outset of the experiment the greater the progress. This relationship was also found in earlier GUME studies.

The general impression is thus one of great variation within methods and between classes, not so between methods. However, since the figure may obscure individual data, and since all computer analyses were made on individual data, we shall procede to them.

Individual progress per method. The progress score for the three teaching methods were analysed with various background variables under statistical control. In an analysis of covariance the choice of

Figure 7: The Progress of the 27 Experimental Classes.



covariates is always a critical question and no rules of thumb exist for choosing. In the present analysis - with Progress as the dependent variable - we believe that the IQ and Pre-test scores are the relevant covariates. The grades are not included, simply because they had been given at a time when the teachers had not had the opportunity to use ary standardized test as a check. The Post-test, although it correlates substantially with the dependent variable, is a pure post-experimental measure and should thus not be used as a covariate. The Standardized test as well as PACT have been used as covariates, which is perhaps somewhat debatable and therefore needs an explanation: Both the Standardized test and PACT might be considered post-experimental measures since they were administered after the GUME investigation was completed. However, it is uncertain to what extent the treatments applied in the experiment have changed the pupils' standing on the Standardized test which is designed so as to correspond to the general objectives of the 3-year course of English in "nellanstadiet" (the pupils are in their third year of English and no standardized test was administered before this one). Likewise, it may be argued that the instructional objectives measured by PACT are too broad for a twelve-lesson series to reach. Besides, PACT was thought to compensate for the fact that the audiolingual aspects of the subject are set aside to a certain extent by the Standardized test.

Thus, two of the four covariates used in the following analyses are completely independent of the treatments and should be relevant for analysis purposes (IQ, Pre-test) whereas two of them (the Standardized test, PACT), although not independent of the treatments, may be defended on the grounds given above. In the table below four separate analyses of covariance are given, each with one of the above-mentioned measures as the covariate. In a fifth analysis the four covariates are weighted together. Thus, in this last analysis, "everything" is held constant.

Table 16 : Analyses of covariance
Dependent variable: PROGRESS
Covariates: IQ, Pre-test, Standardized English Test, PACT and the Weighted Sum of the Four.

Covariates:	Adjusted means				ss'_y			
	Im	Ee	Es	F-ratio	between	within	df	b_w
IQ	16.68	17.44	17.66	.333	94	78520	2/557	.361
Pre-test	17.01	17.34	17.40	.061	16	78739	2/570	.182
Std Engl. Test	17.74	16.74	17.78	.620	132	60020	2/562	.360
PACT	17.38	17.39	17.56	.017	4	65616	2/543	.638
T o t a l	17.99	16.64	17.90	1.086	202	48896	2/524	

ss'_y = adjusted sum of squares in the dependent variable

b_w = the within-groups regression coefficient

Obviously there are no differences between the progress scores for the three teaching strategies. The F-ratios are so low as to make consideration of tendencies among the figures meaningless. Thus the results so far correspond to those obtained in earlier GUME studies: the three treatments, i.e. the teaching strategies Im/Ee/Es, do not produce any significantly different learning effects.

The above table could perhaps be considered "*the* table" of this report, containing information on the main effects in the case of the main dependent variable; as such, the table should perhaps be commented on at greater length. However, we prefer to present all the analyses before discussing the results.

Progress - Interaction.

Although no main effects were found with respect to Progress, it is still possible for interaction effects to exist, i.e. one teaching method may prove superior at one level of ability and another method at another level of ability. Therefore the experimental group was divided into three ability groups (cf p. 57) and the progress scores analysed as in the following table.

Table 17 : Analysis of variance (two-way).
Dependent variable: PROGRESS

Ability level	Teaching Method Im	Ee	Es	Total:
U	18.87	21.24	21.73	20.74
	(54)	(70)	(66)	(190)
M	17.86	18.92	17.83	18.17
	(73)	(60)	(65)	(198)
L	11.51	12.43	13.18	12.47
	(45)	(63)	(65)	(173)
Total:	16.52	17.64	17.60	17.28
	(172)	(193)	(196)	(561)

Source of variation	Sum of squares	df	Variance estimate
Rows (U, M, L)	6423	2	3211
Columns (Im, Ee, Es)	235	2	117
Interaction	158	4	39
Within cells	78599	552	142
Total:	85415	560	

$F_i = .277$ $F_c = .827$ $F_r = 22.557$

The column (i.e. differences between methods) effect is non-significant, thus confirming the results from the preceding analyses. However, the interaction term is also non-significant, indicating that teaching method and ability do not co-vary. Thus the hypothesis that different methods should suit pupils on different levels of ability, is refuted by our data. The row effect (i.e., differences between ability levels) is strongly significant, indicating that pupils of higher intellectual ability learnt more during the experiment than did those of lower ability.

Progress - Main Effects at Two Ability Levels.

There was no interaction between teaching method and intellectual ability. However, table 17 indicates that in the Upper and Lower groups the two E methods tend to give better results. These two ability levels were investigated separately by analysis of covariance to find out whether the differences obtained were significant. In both the analyses the Pre-test scores were used as the covariate.

Table 18 : Analyses of Covariance at Two Levels of Ability (Upper and Lower).
Dependent variable: PROGRESS
Covariate: Pre-test

	Adjusted means				ss'_y			
	Im	Ee	Es	F-ratio	between	within	df	b_w
Upper	18.86	21.16	21.82	1.247	280	20867	2/186	.119
Lower	12.15	12.37	12.80	.044	12	23252	2/169	.171

No treatment effects are discernable when pupils at various ability levels are analysed separately.

To sum up:
With respect to Progress, i.e. learning increment during the experiment, the three methods seem to be of equal capacity. No significant difference was found, nor was there any interaction between method and ability level. Tendencies for the E groups to be better at the upper and lower levels of ability could be explained as chance variation.

Post-test - Main Effects.

In this section two analyses will be presented which are analogous to the first two analyses above. Post-test scores are treated as the dependent variable and Pre-test scores are used as the covariate. Though it is not very likely that the Post-test analyses will give results different from those obtained with Progress as the dependent variable, the analyses should be undertaken in order to disclose whatever information the data may contain.

Table 19 : Analysis of Covariance.
Dependent variable: POST-TEST
Covariate: Pre-test

Adjusted means				ss'_y			
Im	Ee	Es	F-ratio	between	within	df	b_w
68.49	68.83	68.89	1.062	17	78739	2/570	1.182

The within-groups regression coefficient is very high, indicating that a substantial increase in precision is gained by using the Pre-test as a covariate. The adjusted means are, curiously enough, almost identical. No differences whatever exist between the three methods.

Post-test - Interaction.

Again, in order to investigate if any interaction exists, this time between intelligence and achievement on the Post-test, an analysis of variance (two-way classification) was performed. The results are given in table 20 (next page). No interaction is documented in the table.

The tendency for the explicit methods to excel (F-value for columns: 2.664) should be viewed against the background of table 19 i.e. when the pre-test scores are taken into account the differences disappear almost completely.

Table 20 : Analysis of variance (two-way).
Dependent variable: POST-TEST

Ability level	Teaching Method Im	Ee	Es	Total:
U	82.44 (54)	85.28 (71)	84.44 (66)	84.19 (191)
M	64.26 (73)	72.07 (60)	69.94 (65)	68.49 (198)
L	48.20 (45)	53.16 (63)	55.33 (66)	52.70 (174)
Total:	65.77 (172)	70.76 (194)	69.90 (197)	68.94 (563)

Source of variation	Sum of squares	df	Variance estimate
Rows (U, M, L)	90334	2	45167
Columns (Im, Ee, Es)	3106	2	1553
Interaction	738	4	185
Within cells	322980	554	583
Total:		562	

$F_i = .317$ $F_c = 2.664$ $F_r = 77.474$

Progress - the School Class Mean as the Unit of Analysis.

It might be argued that the school class mean is the proper unit of analysis in a study like the present one (see page 43). In the present study this would give only 27 observations, i.e. a great loss of degrees of freedom is made when the analysis moves from the individual to the school class level. However, an analysis of covariance of the school class means on the Post-test was made with the school class means on the Pre-test as the covariate. The result is summarized in table 21.

Table 21 : Analysis of Covariance of School Class Means (N = 27)
Dependent variable: POST-TEST
Covariate: Pre-test

Sources	df	ss_x	sp	ss_y	ss'_y	df	ms'_y
Between	2	63.63	79.63	100.07	.29	2	.15
Within	24	1 907.49	2 384.77	3 241.15	259.68	23	11.29
Total	26	1 971.12	2 464.45	3 341.22	259.97	25	

$$F = {}^{.15}/_{11.29} = .013$$

(Symbols as in Lindquist, 1953)

The F-ratio is almost zero. Thus when the analysis is undertaken at the school class level, every trace of a difference between methods disappears.

To sum up:
The overall impression from the analyses performed thus far is one of non-significant differences between the teaching methods. Nor is there any significant interaction between ability level and method, i.e. no method proves better for pupils on a certain intellectual level. The only strongly significant differences are found between ability groups; pupils of higher intellectual ability score higher and progress more than do pupils of lower ability.

Additional Studies of Progress.

The analyses in the preceding section were all performed on raw scores. However, as was mentioned earlier, two other measures of progress were used (concerning the rationale for using them,

see page 57f). In the case of the first one (actual/possible improvement x 100) an analysis of variance, one-way classification, was performed to find out if the three teaching methods produced different progress. In this analysis the unit of measurement was the school class mean and not the individual score, however. Since this was the case, and since therefore the number of observations is limited, the value for each school class is given in the following table.

Table 22 : School Class Means on the Variable: (actual/possible improvement x 100). N = 27

	Teaching method		
	Im	Ee	Es
	12.09	10.07	11.00
	18.26	16.36	19.32
	22.14	17.97	24.53
	11.75	20.97	20.51
	16.66	15.31	17.29
	20.75	17.27	13.83
	10.21	13.73	16.67
	8.86	25.12	11.04
	18.40	10.62	14.99
N:	9	9	9
$\bar{x}$:	15.46	16.38	16.58

Inspection of the three series of values gives the immediate impression that variation between the three methods is moderate whereas variation between classes within methods is great. The analysis of variance of the results is given in the table below.

Table 23 : Analysis of Variance (one-way classification) of School Class Means on the Variable: (actual/possible improvement x 100).

Source of variation	Sum of sqs	Df	Variance estimate	F-ratio
Between	6.42	2	3.21	
Within	527.30	24	21.97	
Total:	533.72	26		.146

The F-ratio clearly indicates that teaching method does not affect progress measured in this manner.

The second additional measure of progress was: (Post-test-Pre-test/Pre-test) x 100. This measure gives comparatively great credit to progress scores for pupils who had low initial (= Pre-test) scores. As in the case of the preceding measure the school class mean is the unit of analysis. The values for each experimental class is given in the following table.

Table 24 : School Class Means on the Variable: (Post-test-Pre-test/Pre-test) x 100. N = 27

Im	Ee	Es
30.0	28.4	29.2
35.0	37.8	35.9
31.2	26.8	39.5
36.9	32.9	32.0
46.4	25.0	27.6
33.6	36.6	30.7
30.6	31.7	37.0
20.9	45.4	31.3
46.4	32.9	38.2
N: 9	9	9
$\bar{x}$: 33.5	33.2	33.6

The variation in scores between classes within methods is rather great whereas the variation between methods is negligible. In this case no further analysis of the data was undertaken.

To sum up:

The two additional measures of progress, which might perhaps be looked upon as desperate endeavours to find significant differences, gave no information that deviated from that provided by the raw scores.

Thus it seems to make little difference which of the three teaching methods is used. Similarly it seems to make little difference how the progress score is calculated; the results become approximately the same.

Drop-outs.

The drop-outs that will be referred to here are the pupils (N = 43) who were absent from three or more lessons. In order to find out whether the drop-outs deviate in any systematic way from the experimental group, a number of comparisons between the two groups were made. The result of the comparisons are presented in the table below.

Table 25 : Means and Standard Deviations for the Experimental Population and Drop-outs.

	Population (=pupils present 10-12 lessons)			Drop-outs (=pupils absent 3 lessons or more)				
Variable:	N	$\bar{x}$	s	N	$\bar{x}$	s	t	sign
IQ total	564	53.66	9.64	42	53.86	8.82	- .14	
Grades total	573	27.95	7.71	43	26.51	7.77	1.17	
Std test	569	53.48	18.68	41	51.71	16.09	.67	
PACT	550	34.29	8.77	41	33.54	8.71	.53	
Pre-test	575	51.61	20.89	42	49.10	16.61	.93	
Post-test	576	68.67	27.16	42	61.88	24.42	1.73	
Progress	574	17.26	12.32	42	12.79	12.61	2.22	x
Pupil Attit.	529	22.94	4.41	39	22.62	4.60	.42	
Absence	577	.40	.65	42	3.88	1.19	18.66	xx

It seems quite reasonable to assume that pupils with lower grades, intelligence or general knowledge of English would skip the experimental hours more than other pupils. However, the results in the above table indicate that no such selection mechanism lies behind the absence. In fact, no differences are found between the experimental group and the drop-outs in the "pure" background variables (IQ, Grades and Pre-test) or in the others, perhaps not completely unaffected by the treatment (the Standardized test, PACT). The only variable where a significant difference occurs (with the natural exception of Absence) is Progress where the experimental population scores higher. This is a clear indication of a correlation between time spent in class and progress; it pays to be there.

Some Findings in Social Group "0".

As was mentioned above (see p. 60) the 0 stands for cases where the mother (without any mention of her occupation) is responsible for the care of the child. Fifty-one such cases appeared in our population; their distribution on teaching methods was Im: 13, Ee: 15, Es: 23. A X^2 - test shows that this distribution does not deviate significantly from random distribution of cases among methods ($X^2 = 3.30$, $df = 2$, $p > .10$). In order to find out whether this group deviated from the experimental population at large, comparisons were made in a number of variables.

Table 26 : Means and Standard Deviations for the Experimental Population and Social Group "0".

	The Experimental Population			Social Group "0"				
Variable:	N	$\bar{x}$	s	N	$\bar{x}$	s	t	sign
IQ total	564	53.66	9.64	49	51.33	8.39	1.83	
Grades total	573	27.95	7.71	51	24.94	7.58	2.57	x
Std test	569	53.48	18.68	50	49.26	20.23	2.21	x
PACT	550	34.29	8.77	45	31.96	9.08	1.77	
Pre-test	575	51.61	20.89	51	47.84	19.11	1.99	x
Post-test	576	68.67	27.16	51	63.57	26.46	2.32	x
Progress	574	17.26	12.32	51	15.73	12.71	1.01	
Pupil Attit.	529	22.94	4.41	46	23.13	3.91	- .22	

As it appears, there is a clear tendency for this particular "social group" to score lower than the experimental population. Although the data should not be pressed unduly, two conclusions seem possible: either there is an over-representation of social class 3 among the cases under consideration (considering the correlation between social class and achievement), or there is a connection between a mother as the sole guardian and low scores on the part of the child.

Findings Related to Course Choice.

As was mentioned above (see p. 61) the pupils, in February, made their choice as regards course in English for grades 7 through 9. It was considered interesting to investigate whether this choice was associated with the pupils' standing on various background variables. Some information relating to this question will be given in the tables and figures below.

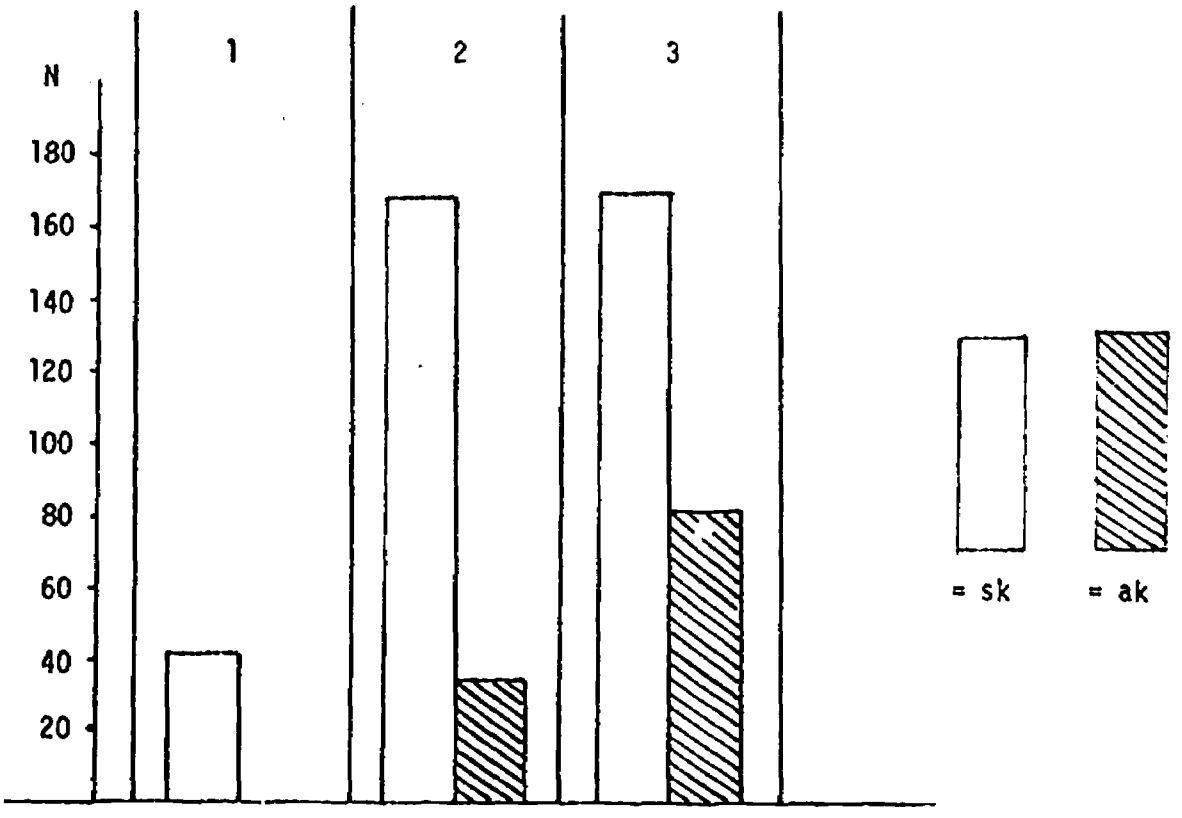
Table 27: Distribution of Social Class in Relation to Course Choice (Absolute numbers to the left, percentages to the right).

	Social Class				Social Class			
	1	2	3	Total	1	2	3	Total
Sk	41	169	170	380	10.8	44.5	44.7	100.0
Ak	0	35	81	116	0	30.2	69.8	100.0
Total	41	204	251	496	8.3	41.1	50.6	100.0

It is apparent from these figures that the choice of course in English is associated with social class. A X^2-test gave a X^2-value of 26.85 ($df = 2$; $p < .001$); there is thus a strong reason for rejecting the null hypothesis of independence between social class and course choice.

The figure below is intended to visualize the relation between social class and course choice.

Figure 8: Distribution of sk/ak Choices per Social Class.
(N_{sk} = 380) (N_{ak} = 116)



A comparison was also made between the two "categories" of pupils in some other variables. The results are given in the following table.

Table 28 : Means and Standard Deviations for Presumptive sk and ak Pupils.

	sk			ak				
Variable:	N	$\bar{x}$	s	N	$\bar{x}$	s	t	sign
IQ total	428	55.50	9.22	133	47.76	8.49	7.10	xx
Grades total	428	30.69	6.46	142	19.75	4.89	12.88	xx
Std test	428	59.07	16.76	139	36.20	12.96	16.57	xx
PACT	417	36.18	7.98	131	28.37	8.33	7.37	xx
Pre-test	431	56.62	20.80	141	36.53	12.13	14.14	xx
Post-test	431	75.96	26.02	142	46.79	16.94	17.90	xx
Progress	430	19.50	11.96	141	10.43	10.63	7.43	xx
Pupil Attit.	397	23.18	4.43	129	22.25	4.34	1.21	

The differences between sk and ak are highly significant in all cases except Pupil Attitude. These group differences might be taken to indicate that, in general, the pupils' choices are realistically made with respect to their grades, intellectual standing, and knowledge of English. However, since the choice is of great interest on the individual level, we shall present the sk and ak distributions on the IQ test and two linguistic variables, namely the Standardized test and PACT (see next page).

One salient feature of all three figures is that the distributions are more or less completely overlapping. This phenomenon was discussed in our earlier reports (see, for instance, Levin, 1969, pp 68-70), although it then referred to actual courses and not, as in the present case, courses chosen for the next year. The tendency towards overlap is still more pronounced in this study.

The distributions seem to warrant the following reflexion on the fact that in English, from grade 7 and onwards, the pupils are divided into two courses:

If it were assumed that the pupils' general intellectual ability and/or knowledge of English up to grade 6 should guide their course choice, then a fairly large number of pupils seem to make ill-advised choices; it is notable that pupils with relatively high intelligence and language test scores choose the less advanced course and, correspondingly, that pupils of low intelligence and language test scores take the more advanced course. One might argue, of course, that the former group of pupils, despite their capacity, have chosen the easier course because of little or no interest in learning English. However, if we consider the information given above on the relation between social class belongingness and course choice, it is difficult to escape the suspicion that sociological factors are decisive for many pupils.

If this problem is looked upon from the individualization point of view, it becomes evident that although the teaching in sk (according to the higher mean there) may proceed at a relatively higher speed than in ak, the *variation* in general ability and proficiency between pupils is about the same in sk as in the total group. Thus, on the average the need for individualization would be the same (in sk) whether the courses are kept apart or not. The effect of putting the two courses together would be - still from the sk teacher's point of

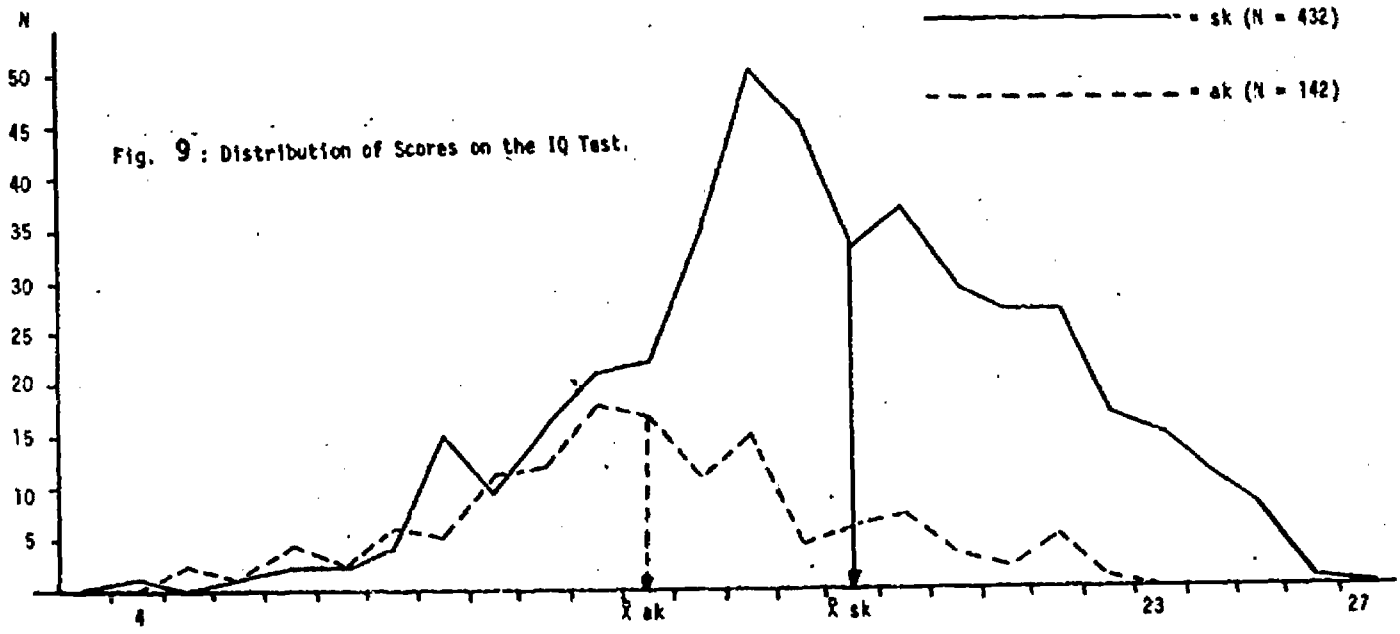


Fig. 9 : Distribution of Scores on the IQ Test.

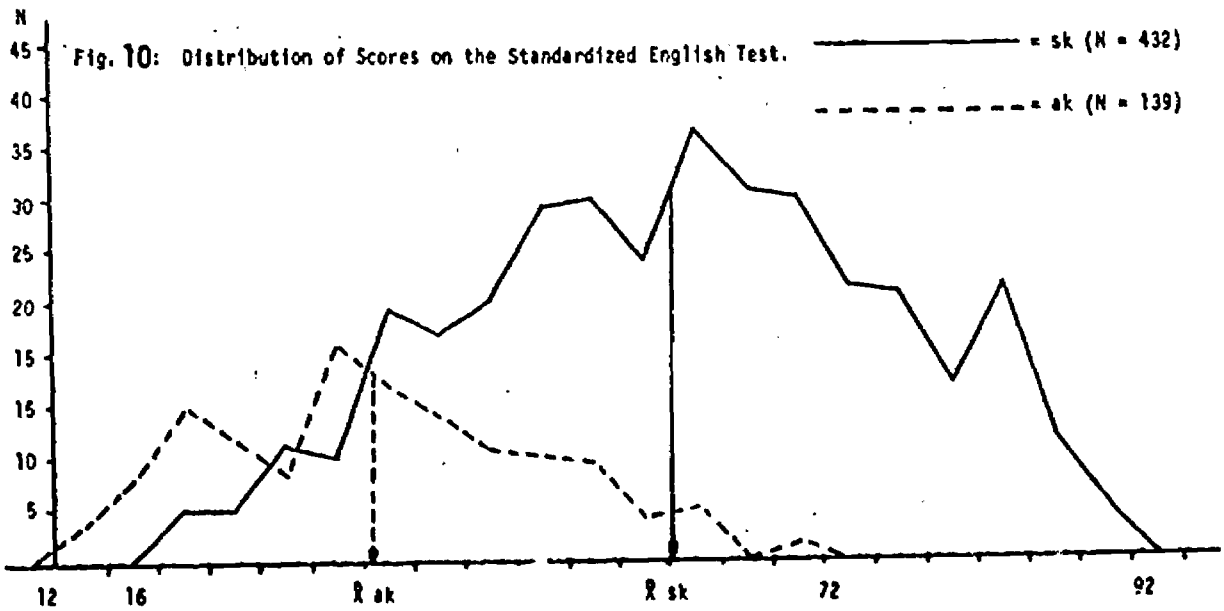


Fig. 10: Distribution of Scores on the Standardized English Test.

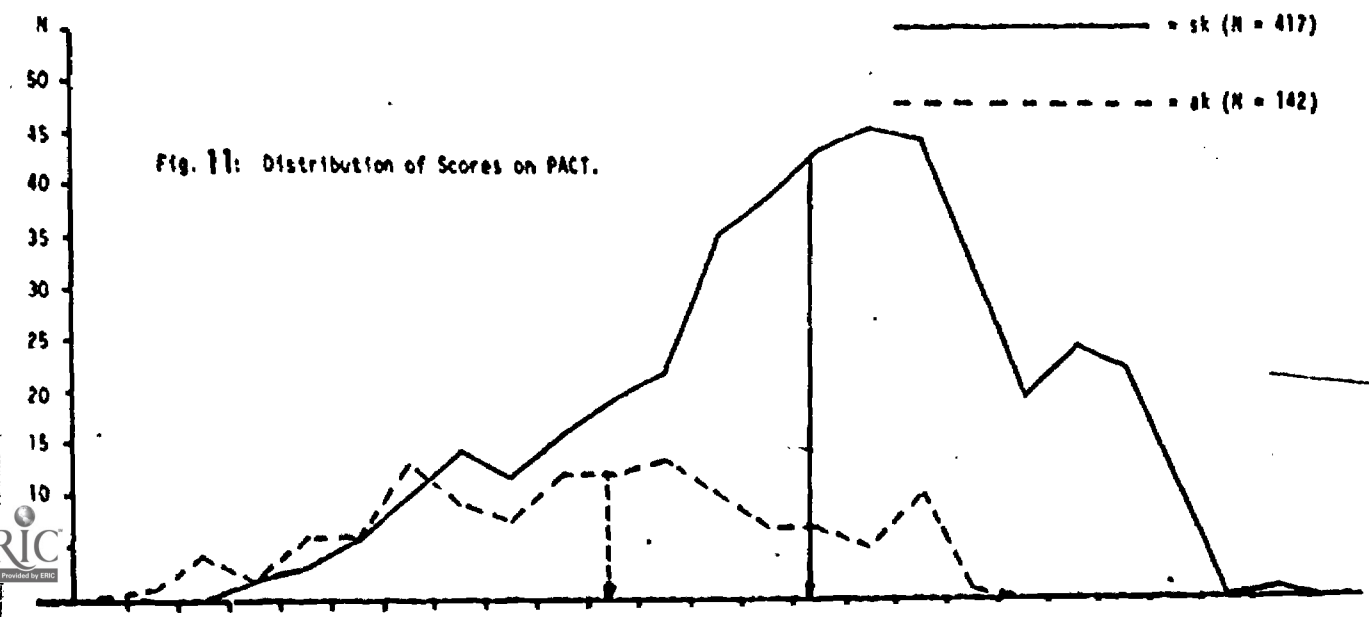


Fig. 11: Distribution of Scores on PACT.

view - that the class would consist of a few more slow learners; again, the *variation* between the bottom and top pupil would be about the same. Looking at the problem from the "ak angle", it is very probable that a number of pupils in this group would profit from being taught together with those of sk.

All in all, as the course choice functions today with a number of factors other than ability and proficiency in English strongly influencing the choice, there seems to be little justification for keeping the two courses separate. What effect a fusion of the two courses would have on discipline and atmosphere in the classroom is an interesting problem but outside the scope of the present project and not for us to discuss.

Results on the Different Part Tests.

The results on the seven parts of the test are given in table 29.

The most difficult test seems to be *part* G, in which the pupils were asked to make negative sentences. They managed only 11.3 % of the items in the pre-test and increased their scores only 5.2 % on this test. It seems probable that many of the pupils did not understand what they were supposed to do, since most of them undoubtedly knew more of this construction than the results seem to indicate. The easiest test on the other hand seems to be *part* D, where more than half the number of items (58.9 %) were correct in the pre-test and 73.8 % in the post-test. In spite of the high figure in the pre-test there was ample room for progress which also became the second highest (14.9 %).

In constructing and trying out the test we aimed at a rate of correct answers in the pre-test of about 30-35 %, and the figure for the pre-test, 32.3 %, is thus very satisfactory. The test was very difficult and the post-test figure, 42.9 %, is perhaps a little lower than expected. It also indicates that the test can, in all likelihood, be used with good discriminating power in the 7th, and possibly the 8th, form also.

Part A is undoubtedly the easiest in one respect: they have been working with this problem for almost three years. It is quite clear from the figures that our pupils in leaving "mellanstadiet", after three years of English with a total of 11 hours per week, do not know how to use the s-form of verbs in the third person singular and that a lot more practise is needed.

Parts B *and* E test the pupils' ability to make questions, primarily questions with do/does/did. They result in exactly the same kind of phrases but the stimuli are different: in test B the pupils are told to "ask me if..." and in E they have a sentence and are told to ask a question with the same verb but a different object. In the pre-test there are no differences between the tests, the percentages of correct answers are 13.5 and 14.2 respectively. In the post-test, however, there is a noticeable difference: 28.9 and 25.2, which indicates that the more mechanical way of testing used in B is easier; test E probably requires a larger amount of intellectual abilities.

Part C consisted of various items, the reason being that it is difficult to test prep + ing-form by itself since it tends to give a

Table 29 : Results on the Parts of the Tests per Method.

	Im		Ee		Es			
Pre-test:	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	% x)	(max. score)
A	3.17	1.97	3.42	1.79	3.21	1.84	32.7	(10)
B	1.92	2.57	2.18	2.51	1.96	2.50	13.5	(15)
C	16.92	6.25	17.63	6.46	17.29	6.09	38.4	(45)
D	11.19	4.81	12.16	4.67	11.90	4.68	58.9	(20)
E	1.91	2.61	2.33	2.70	2.14	2.43	14.2	(15)
F	12.62	5.77	13.77	6.62	13.76	6.04	33.5	(40)
G	1.50	2.55	1.65	2.32	1.91	2.60	11.3	(15)
Totals	49.24	21.46	53.14	21.25	52.27	19.91	32.3	(160)
Post-test:								
A	4.36	2.19	4.23	2.17	4.65	2.02	44.2	
B	3.73	3.67	4.63	4.25	4.57	4.04	28.9	
C	20.98	7.38	21.73	7.66	21.17	7.72	47.3	
D	14.24	4.88	15.12	4.85	14.86	4.57	73.8	
E	3.29	3.68	3.97	4.08	4.04	3.86	25.2	
F	17.02	7.44	18.43	8.10	17.64	7.44	44.3	
G	2.04	3.10	2.68	3.28	2.65	3.37	16.5	
Totals	65.35	25.70	70.79	28.14	69.58	27.30	42.9	
Progress:								
A	1.20	2.09	.81	2.02	1.44	2.11	11.5	
B	1.82	2.68	2.44	3.28	2.61	3.00	15.4	
C	4.06	5.72	4.11	5.49	3.88	5.41	8.9	
D	3.05	3.70	2.96	3.01	2.96	3.35	14.9	
E	1.38	2.31	1.64	2.75	1.91	2.63	11.0	
F	4.46	4.95	4.66	4.91	3.88	5.48	10.8	
G	.54	2.05	1.04	2.28	.75	2.26	5.2	
Totals	16.52	11.95	17.64	12.35	17.54	12.65	10.6	

Key to the tests:

A: answers to questions; tests the s-form
B: make questions; tests the correct use of the do-construction
C: four-choice test: tests prep+ing-form, the continuous tense, the s-form
D: position of adverbs of time, tests correct placing of these
E: make questions: tests use of do-construction (same as B but different stimuli)
F: six-choice test of the some-any problem
G: make negative sentences: tests the use of the do-construction in negative sentences

x) The % figures refer to the total mean for all pupils in relation to the possible number of items per part test; the progress figures are the differences in per cent for the post- and pre-tests.

result of "all or nothing". Thirty of the 45 items were investigated as two different tests and will be discussed below as "critical items".

Tests D and E have already been commented on.

Part F, which was a multiple-choice test, was the only one testing the some-any dichotomy which was one of the major parts of the lessons. The number of correct answers here, 13.4 for all pupils, is 33.5 % of the total possible. This increases in the post-test to 44.3 %. A special study has been made of those items in which the use of 'some' and 'any' did not follow the basic rules, e.g. Would you like some coffee ('some' in questions) and Anybody can do that ('any' in ordinary statements).

Test G has been commented on above.

Method differences. We were also interested in studying differences between the methods on the various tests since some of the structures dealt with were new and some well-known or at least practised in class for some time.

On the pre-test all differences between the methods are small. One tendency is noticeable, however: the Im group scores lowest on all the seven parts, and Ee higher than Es on all except G.

On the post-test the situation is almost identical, except that Im has passed Ee on test A and that Ee and Es (which are very close) have changed places on some tests. There are no differences, however, which are large enough to warrant special attention, and the results on the parts are thus identical to those for the whole test.

In studying progress scores we notice the large standard deviations, especially large on test G compared to the low means. These figures indicate that many pupils scored 0 and that there is a marked positive skew in the distribution.

"Critical Items".

The so-called critical items were investigated in order to bring about an analysis of certain items contained in larger tests. They include a total of 51 of the 160 items of the test.

Test A included those 6 items in test A that required an -s (the remaining 4 were included as distractors). The mean is just a little below that for the whole test and the percentage is much higher

(46.3 as compared to 32.7 on test A as a whole). This seems to indicate that the pupils missed many of the easy items without an -s and that, knowing what the test was about, they tended to use "too many s-es". There are no differences between methods.

Test CA includes all examples in part C also requiring the s-form. The percentage correct here is the same as for test A and slightly below the mean for the whole test. This again underscores the impression that this high-frequency structure (third person singular present tense) is very poorly mastered after three years of study. Even in the post-test, after an intensive period of practise and commenting, only 40.5 % of the items were correct. There are no noticeable differences between the methods; the progress scores (1.03, .99 and 1.02) are all-but identical whether the pupils were given explanations or not. Even in this little seemingly simple test the result of the whole study is mirrored quite completely.

Test CB is really a test in its own right, mixed into a longer test for reasons mentioned above. This structure (prep + ing-form) is completely new to the pupils and is normally dealt with only at higher stages. It is also in sharp contrast to Swedish usage, and therefore this is one of the points at which differences ought to come out most clearly.

In the pre-test the pupils got almost 8 out of the 19 items correct, which is equal to 40.3 per cent. It should be borne in mind, however, that this was a 4-choice test, and after correction for guessing there are only about 4 correct items left. Even this figure is quite high for a completely unknown structure. There may be two reasons for this: the pupils hear a lot of English on TV and radio and some phrases might have been known to them for this reason, and, secondly, since there were four choices they might have ruled out the other three (infinitive, s-form, present continuous) as impossible and thus taken the unknown - and correct - structure.

The total progress (11.8 %) is almost the same as that for the whole test (10.6 %) and thus neither low nor particularly great. There are no differences between methods; and it is worth noticing that the Es progress is the smallest, although the difference is not significant.

Table 30 : So-Called Critical Items per Method.

	Im		Ee		Es				Total
	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	max	%	$\bar{x}$
Pre-test	17.99	6.33	19.01	7.08	18.85	6.86	51	36.5	18.64
A	2.65	1.55	2.84	1.46	2.83	1.57	6	46.3	2.78
CA	3.29	1.97	3.66	1.97	3.38	1.86	11	31.4	3.45
CB	7.41	3.17	7.74	3.49	7.77	3.51	19	40.3	7.65
F	4.63	2.32	4.78	2.69	4.87	2.48	15	31.7	4.76
Post-test	23.80	8.10	24.34	8.90	24.09	7.93	51	47.2	24.09
A	3.55	1.67	3.37	1.63	3.82	1.57	6	59.7	3.58
CA	4.32	2.35	4.65	2.31	4.40	2.35	11	40.5	4.46
CB	9.77	3.92	10.10	4.65	9.79	4.07	19	52.1	9.89
F	6.17	2.86	6.22	3.24	6.08	2.80	15	41.1	6.16
Progress	5.83	6.06	5.33	6.30	5.25	5.34		10.7	5.45
A	.88	1.80	.54	1.70	.99	1.73		13.4	.80
CA	1.03	2.35	.99	2.28	1.02	2.19		9.1	1.01
CB	2.38	3.42	2.36	4.05	2.03	3.27		11.8	2.24
F	1.54	2.53	1.44	2.50	1.21	2.47		9.4	1.40

Key: A: the s-form, all examples in test A requiring answers with verbs in -s
CA: the s-form: all examples in test C requiring verbs in -s
CB: prep + ing-form: all examples in test C requiring an ing-form after prepositions plus four examples with to + infinitive as distractors
F: 'some-any': all examples in test F where the use of 'some' and 'any' (and their compounds) differs from the basic rules ('some' in questions etc.)

NOTE: Parts A and CA above should be compared to part A in the test as a whole.
Part CB is to be considered a part test in its own right, comparable to the other part tests.
Part F is the only part here which really consists of "critical" items, i.e. items where a special difficulty exists, one that might cause differences in the results between methods different from the overall results.

Test F finally consists of 15 items which are really "critical" in the sense that they deviate from the norm. They might be said to test the "feeling" for the use of 'some' and 'any' and not knowledge of any rules. The result here is the same as in all other cases: about

one third correct on the pre-test, about 10 % progress, no differences between methods. And again we notice that Es makes the smallest progress.

To sum up:

All figures on the parts point in the same direction as those for the whole test, and an investigation of the parts in view of the fact that they test different structures of which the pupils have had different experience yields no interesting results. It seems that progress is about equal over an intensive period of drilling and practising whether the structure being practised has been taught for years or is completely new, and this seems true irrespective of method used.

CORRELATION STUDIES

All variables used in the project have been inter-correlated and will be discussed here. The general impression of the correlation tables (see, for example, table 33) is that most figures are relatively or very high and even. This first impression seems to bear out the finding of, for example, Carroll (1958 , p. 16) that it is hard to find a clear factorial pattern in linguistic competence. It is not possible to find different factors at work, resulting in different correlations, in the listening tests, pronunciation test, grammar tests, reading comprehension test etc.

Background Variables.

In table 31 correlations are given for a number of variables for the whole pupil population irrespective of method and intelligence level.

Social class correlates around .20 with IQ as well as with measures of scholastic aptitude and proficiency in English. This well-established fact, which is brought out in all similar studies, is interesting but not surprising.

Pupil attitude generally shows low correlations. The highest correlation is with the post-test, which is an indication that the more interested pupils have done better in the project, or perhaps rather: that those who felt they had made progress had a more favourable attitude towards the project when the attitude test was given. There are also positive correlations with PACT and the Standardized test, which were both given after the project. This might indicate that those who were positive after the project were more motivated to do their best on those tests. Interestingly enough there is also a slight correlation with grades in English.

Intelligence correlates significantly with all factors except attitude. Grades in English correlate higher with the Verbal IQ than with the IQ total, but the reverse is true for Maths. Spatial IQ has the lowest figures throughout, correlations with grades in Maths being the only one coming close to .40. Both Verbal IQ and the IQ total correlate significantly higher with the pre- and post-tests than with progress,

Table 31: Intercorrelations between Main Variables. (N = 577)

		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1.	Social class	.200	.164	.043	.176	.175	.226	.206	.233	.257	.174	.189	.174	.229	.238	.214	.219	.126	.104
2.	Verbal IQ		.380	.264	.702	.609	.564	.478	.624	.636	.598	.576	.533	.684	.523	.586	.604	.105	.345
3.	Inductive IQ			.446	.799	.420	.372	.554	.508	.404	.391	.374	.411	.461	.370	.408	.418	.077	.238
4.	Spatial IQ				.766	.218	.181	.397	.301	.248	.211	.240	.237	.272	.203	.220	.208	.053	.069
5.	IQ,Total					.529	.474	.618	.614	.551	.517	.509	.505	.607	.471	.520	.527	.095	.282
6.	Grades Sw.						.807	.608	.900	.600	.652	.586	.551	.729	.469	.633	.660	.160	.380
7.	Grades Eng.							.610	.911	.637	.709	.664	.668	.784	.545	.697	.735	.206	.448
8.	Grades Math.								.833	.503	.540	.489	.515	.599	.391	.492	.517	.059	.316
9.	Grades total									.656	.715	.656	.691	.797	.530	.682	.719	.171	.434
10.	Stand. EL										.687	.637	.616	.848	.631	.729	.753	.162	.457
11.	Stand. EM											.577	.724	.886	.604	.750	.784	.143	.466
12.	Stand. EA												.618	.824	.660	.676	.734	.228	.477
13.	Stand. EU													.853	.567	.689	.734	.246	.452
14.	Stand. total														.717	.832	.882	.226	.542
15.	PACT															.667	.718	.246	.454
16.	Pre-test																.902	.181	.307
17.	Post-test																	.264	.588
18.	Pupil attit.																		.258
19.	Progress																		

although even the latter are significantly higher than 0. This indicates that both pupils of low and those of medium and high intelligence have progressed in the project but that those in the upper echalons have made slightly greater progress.

Grades, the three separate measures as well as the total, correlate significantly with all other factors. These figures are mostly significantly higher than those for intelligence. If progress is taken to mean success in English, then the teachers' subjective grades in English are a better prognostic instrument than IQ, be it the total or only the Verbal part. Even grades in Maths is as good as the IQ total. An IQ testing is a quicker and no doubt more reliable measure to use if one were to give a prognosis for an unknown pupil, but this testing cannot compare in value with the evaluation by a teacher who has known the pupil for almost three years, if this measure is available.

Again we also notice that the pre- and post-tests correlate higher than does progress. Teachers' grades in English, which have been given after 2.5 years of instruction but with no outside help in the form of standardized tests, thus seem to be the best predictor of success in the study of English. The figures for the Standardized test are higher still, but then it should be borne in mind that they are influenced by the same teaching as influenced the achievement test. In giving a prognosis of success in language studies, for example if parents want advice whether the pupil should take the more advanced course in grade 7 and a second language, a composite measure consisting of grades, results on the standardized test and the Verbal IQ, would in all likelihood be the best possible at the moment.

The Standardized test and its parts correlate well with teachers' grades (.78 with English) and with the project tests.

PACT, the listening comprehension test, correlates well with grades and other tests, but the difference between grades in English (.55) and with the Standardized test (.72) seems to indicate that listening comprehension might be a slightly neglected factor in giving grades. The high correlation with the all-written pre- and post-tests (.67 and .72) show that even a test which is all written gives a good

overall evaluation of the pupil (Cf what was said above at the beginning of this section about the lack of clear factorial patterns).

The correlation between the pre-test and the post-test (.90) is the highest in the whole matrix, which might be taken as an indication that our Achievement test has a high reliability.

We shall give a word of comment on a point which is self-evident to those of our readers who are well versed in statistical matters but might not be so to others with lesser statistical training. The correlations between the pre- and post-tests on the one hand and almost any other factor on the other are higher than the corresponding correlations between progress and these same factors. The correlations, for example, between IQ and the pre- and post-tests are .68 and .72 respectively, but between IQ and progress .43. The explanation is that correlations are dependent on the size of the standard deviations; compare the post-test, for example, ($\bar{x}$ 68.67, s 27.16) and progress ($\bar{x}$ 17.26, s 12.32). This is not surprising since progress is the difference between the pre-test and the post-test; the difference between two figures, of course, is smaller than the figures themselves (It is also normal for the standard deviation to be lower when the mean is lower).

Correlations with Course Choice.

Correlations have also been calculated between the pupils' choice of course in grade 7 (ak, the easier one, and sk, the more difficult one, which is taken by roughly 75 % of the group) and certain other variables. They are as given in table 32 (see next page).

All these figures, except for social class and attitude, are highly significant. What is interesting, and somewhat disconcerting, is the fact that the correlation with grades, e.g. in English (.60), is much greater than the one with knowledge of English as measured by the Pre-test (.41), which is an indication that there is a tendency for those who have failed, for one reason or another, to gain good grades, to take the easier course whether their skills warrant it or not.

Table 32: Correlations between the Pupils' Course Choice for Grade 7 and Certain Other Variables.

	Course choice for grade 7
Social class	-.08
IQ Verbal	.40
Inductive	.28
Spatial	.13
Total	.35
Grades, Swedish	.55
English	.60
Maths	.49
Total	.61
Stand. Test, EL	.42
EM	.48
EA	.46
EU	.43
Total	.53
PACT	.38
Pre-test	.41
Post-test	.46
Progress	.32
Attitude	.09

Pre-test and Post-test Correlations.

In correlating the various parts of the pre- and post-tests with grades (table 33), we find that test A is lower than the others, whereas all the others seem to have roughly the same figures (between .50 and .60). They correlate almost exactly the same with the Standardized test and its parts. The correlation between the post-test total and the Standardized test total is .88, which means that about 77 % of the variance is explained by common factors. PACT also correlates well and these high correlations are an indication that the achievement test has high validity.

The parts of the achievement test correlate well with the Verbal IQ test (the total being about .60), less with the Inductive part (.40) and very little with the Spatial test (about .20).

Table 33: Pre- and Post-test Correlations with Certain Other Variables.

		BETYG				STD-test						IQ			
		Sw	Eng	Ma	Tot.	EL	EM	EA	EU	Tot.	PACT	Verb.	Ind.	Spat.	Tot.
Pre-	A	.331	.321	.264	.344	.287	.293	.284	.303	.340	.241	.199	.240	.105	.231
test	B	.461	.500	.356	.487	.515	.550	.495	.537	.613	.421	.393	.300	.140	.354
	C	.500	.557	.414	.552	.577	.607	.514	.554	.659	.537	.425	.311	.173	.393
	D	.538	.616	.420	.594	.607	.663	.641	.606	.737	.623	.559	.329	.200	.466
	E	.529	.557	.386	.548	.585	.591	.512	.567	.659	.479	.466	.302	.189	.409
	F	.524	.575	.377	.553	.648	.623	.582	.534	.697	.607	.538	.371	.182	.465
	G	.472	.534	.391	.520	.559	.571	.469	.556	.630	.457	.429	.308	.176	.389
	Total	.632	.697	.492	.682	.729	.750	.676	.689	.832	.667	.585	.408	.220	.520
Post-	A	.449	.459	.344	.472	.438	.457	.381	.452	.505	.360	.306	.250	.068	.263
test	B	.529	.581	.436	.579	.576	.622	.546	.594	.684	.533	.435	.334	.121	.381
	C	.546	.603	.444	.600	.668	.660	.613	.640	.753	.630	.499	.368	.191	.458
	D	.552	.630	.427	.608	.558	.633	.668	.591	.718	.640	.534	.350	.173	.452
	E	.556	.610	.420	.594	.632	.670	.577	.618	.731	.525	.505	.354	.154	.433
	F	.550	.651	.409	.607	.706	.684	.675	.607	.781	.667	.566	.324	.182	.458
	G	.535	.567	.449	.578	.610	.626	.532	.587	.688	.516	.453	.394	.205	.453
	Total	.660	.735	.517	.719	.763	.784	.734	.736	.882	.718	.604	.418	.208	.527

Table 34: Pre- and Post-test Intercorrelations.

		Pre-test							Post-test							
		B	C	D	E	F	G	Tot.	A	B	C	D	E	F	G	Tot.
Pre-test	A	.364	.374	.246	.353	.295	.407	.481	.462	.400	.332	.241	.363	.234	.425	.392
	B		.538	.472	.703	.521	.669	.743	.413	.661	.620	.405	.671	.525	.650	.692
	C			.519	.585	.597	.597	.836	.435	.586	.697	.430	.599	.578	.621	.708
	D				.518	.610	.493	.765	.374	.536	.582	.751	.579	.667	.509	.732
	E					.606	.712	.798	.482	.653	.646	.446	.753	.616	.672	.750
	F						.581	.847	.420	.568	.625	.532	.641	.746	.575	.752
	G							.788	.507	.628	.644	.417	.667	.571	.738	.726
Total									.550	.737	.787	.632	.787	.776	.760	.902
Post-test	A									.476	.516	.342	.449	.387	.503	.581
	B										.661	.501	.749	.599	.688	.812
	C											.561	.690	.667	.688	.882
	D												.514	.632	.456	.738
	E													.664	.736	.844
	F														.590	.813
	G															.803

In table 34 we see how the various parts of the pre- and post-tests correlate with each other. On the whole, figures for the post-test are higher, which is probably explained by the fact that its means are higher with an accompanying increase in standard deviations, which strongly influences the correlations (cf above). This also explains why tests C and F have the highest correlations with the test totals (.84 and .85 for the pre-test, .88 and .86 for the post-test). These tests explain most of the variance, and they could be given alone to yield, in a very short time, a fairly reliable overall picture of what the pupils know. Test A has the lowest figures throughout, which is partly explained by its relatively low reliability (.52).

Progress Correlations.

Table 35 shows various Progress correlations. We notice here that the progress total correlates significantly with all parts of the pre-test except the first one, and that the correlation with the pre-test total is .31, which is an indication that the better pupils have made better progress than the poorer pupils although only a small part of the variance in progress is explained by the pre-test. Only one pre-test column has nothing but negative figures, that for progress on part D, the position of adverbs. Almost all are non-significant, however, but the test itself correlates with the progress made on it -.34, which definitely means that those who scored poorly on the pre-test learnt most about this.

The correlations between progress and the post-test are higher throughout than those with the pre-test. The post-test and progress totals correlate no less than .688. The progress total correlates well with all parts of the post-test, part A being the lowest with .366. Progress on part D still correlates negatively with all tests (except part D and the total), all being non-significant, however.

These negative correlations on test D need a special word of comment. As table 29 on page 88 shows, the means of the various part tests vary between 11.3 % and 38.4 % of the total possible on the pre-test, except in the case of test D, where this figure is 58.9 %. On the post-test it has risen to 73.8 %. In this test there was thus much less room for progress for the better pupils (as a matter of fact 2.1 % had all 20 items correct in the pre-test, 12.5 % had 18, 19 or

Table 35: Progress Correlations.

		Progress							
		A	B	C	D	E	F	G	Tot.
Pre-test	A	-.423	.228	.032	-.003	.193	.004	.170	.073
	B	.097	.043	.241	-.091	.307	.163	.206	.287
	C	.110	.330	-.177	-.121	.316	.150	.244	.160
	D	.163	.318	.210	-.342	.352	.267	.197	.323
	E	.176	.281	.223	-.097	.132	.197	.191	.318
	F	.165	.321	.181	-.103	.358	-.084	.195	.237
	G	.154	.276	.207	-.103	.292	.159	-.037	.285
	Total	.131	.359	.132	-.181	.386	.141	.235	.307
Post-test	A	.609	.288	.216	-.041	.194	.082	.171	.366
	B	.129	.778	.243	-.043	.474	.219	.309	.559
	C	.231	.361	.583	-.023	.393	.251	.291	.626
	D	.135	.329	.283	.363	.327	.308	.204	.559
	E	.134	.435	.269	-.086	.752	.227	.336	.541
	F	.186	.365	.261	-.040	.382	.601	.236	.596
	G	.133	.372	.241	-.070	.436	.195	.647	.493
	Total	.243	.510	.406	.018	.518	.388	.379	.688.
Progress	A		.091	.193	-.039	.027	.079	.022	.307
	B			.122	.019	.374	.156	.239	.506
	C				.105	.182	.173	.123	.673
	D					-.032	.062	.012	.337
	E						.144	.315	.495
	F							.108	.608
	G								.412

20 correct). The less gifted pupils have come closer (69 pupils, 12.0 %, had 1-5 correct on the pre-test; 39, 6.8 %, on the post-test), and as the standard deviation figures indicate the group is more homogeneous on the post-test; in spite of the fact that the number of points has increased with about 3, the standard deviation is almost exactly the same.

The highly significant figures between the pre- and post-tests and progress should be compared to those obtained in the GUME 1 and 2 follow-up studies, discussed on pages 118-127 below.

In the progress-progress correlations we notice that especially parts C and F carry a heavy load in the total, these also being the longest tests. Again we notice that tests A and D stand out as having the lowest correlations.

ATTITUDES

Pupils' Attitudes to the Project (General).

The overall means (see table 36a) indicate that the Ee pupils are more positive to the project than the others, and, perhaps somewhat surprisingly, that the Es pupils are most critical.

The only two questions which have a mean in all groups below 3.0 are numbers 4 and 11, which indicates that both the lessons on the whole and the oral drills were a little more boring than fun. It is probably a common experience that pupils are reluctant to admit that they like school. It could also be mentioned that the means for these two questions are for boys 2.74 (s 1.12) and 2.60 (1.03) respectively, and for girls 2.96 (.93) and 2.88 (1.04) respectively. Girls are thus more willing to admit that they like school. Most of these means are just below 3.00, except in Es where 2.64 and 2.69 are markedly lower than in the other groups.

The difference here between the methods is largely due to one class which showed a very negative attitude throughout the whole project and whose means for questions 4 and 11 were 1.64 (.79) and 2.27 (1.20) respectively.

The answers to question 3 shows that the pupils felt that they learnt more or less as usual, not very much and not very little. Pupils in Ee have felt that they learnt a little more than the others.

The most positive answers have been given to question 5: Did you understand what you were doing? 82 % of them feel that they understood this always or almost always. Four pupils think they never understood what they were doing! Here the Im pupils who did not get any explanations or comments at all have the highest mean, 4.14! Questions 9 and 10 dealt with the four-phase drills: whether they felt they had learnt to speak and learnt grammar from them. There is no difference in the slightly positive means of question 9, but the differences in number 10 are interesting: the Im pupils *feel* they learnt less grammar than the others. The results on the achievement test as reported elsewhere show that this was not so. The explanation of the difference here (2.96 in Im, 3.38 in Ee, a difference of no less than .42) is, no doubt, the fact that the Im pupils had no explanations and thus did not

Table 36a: Pupils' Attitudes (Explanations excluded).

Question	Frequencies 1	2	3	4	5	No answer	Means and standard deviations: Total (N=529) $\bar{x}$	s	Im (N=167) $\bar{x}$	s	Ee (N=172) $\bar{x}$	s	Es (N=188) $\bar{x}$	s
3	46	79	205	181	18	48	3.09	.98	3.07	.94	3.23	.96	2.97	1.03
4	58	132	178	148	13	48	2.86	1.02	2.99	.95	2.98	1.03	2.64	1.05
5	4	14	76	290	145	48	4.05	.77	4.14	.69	4.03	.83	4.01	.77
9	24	64	221	187	31	50	3.26	.91	3.31	.86	3.37	.88	3.12	.97
10	31	86	212	161	37	50	3.17	.98	2.96	1.02	3.38	.90	3.14	.97
11	65	158	164	126	16	48	2.75	1.04	2.77	1.02	2.81	1.00	2.69	1.10
12	8	24	154	218	125	48	3.81	.90	3.81	.85	3.77	.92	3.84	.92
					Total mean:		3.28		3.29		3.37		3.20	

3 During project lessons I learnt *very little - very much.*

4 Project lessons were *very boring - great fun.*

5 In doing oral and written exercises I understood what to do *never - always.*

9 From the four-phase drills I learnt to speak English *very poorly - very well.*

10 From the four-phase drills I learnt English grammar *very poorly - very well.*

11 The four-phase drills were *very boring - great fun.*

12 The four-phase drills were *very difficult - very easy.*

know what they were learning, did not know that it was grammar. One class, whose teacher was very negative to the Im method and personally preferred Es, had a mean of 2.21 (1.23), with a distribution on the five alternatives 8-3-4-4-0, i.e. eight felt they learnt "very little" and none "very much". It seems likely that the attitude of the teacher had influenced the pupils, not necessarily the teacher's attitude during the project but the method of teaching with a lot of comments previous to the project.

The four-phase drill, finally, was felt to be easy to do, as number 12 shows. No differences between methods, and a strongly positive skew in the answers.

To sum up:
On the whole the attitude of the pupils to the project is leaning towards the positive. The only two differences between methods - slightly more negative attitude to the lessons on the whole in the Es group, and a feeling of learning less grammar in the Im group - may probably both be explained by atypical classes in the groups.

Pupils' Attitudes to the Explanations.

Questions 6,7, and 8 of the Pupils' Questionnaire concerned the explanations given or not given. Question 6 was included to check whether the pupils were aware that they had got any explanations at all. Number 7 was meant for explicit groups only and number 8 for implicit groups only.

Number 6: 41 Im pupils thought that they had got explanations; this is almost 25 %. This seems to indicate that they felt pretty sure what they were supposed to learn and that they had not detected that in fact no theoretical explanations had been given. Even in the class mentioned above whose teacher usually gave explanations and therefore was so dubious about the value of the Im method, there were 4 out of 24 who thought they had had explanations.

In the explicit groups there were 23 and 11 respectively (15 % and 6 %) who thought they had not had any explanations. The figure for Ee is perhaps not so surprising, but that 11 Es pupils never noticed that the teachers on the tape sometimes spoke Swedish is interesting.

Question 7 A: This question, as it turned out, was too complicated; we wanted Ee pupils to say whether they would have preferred to have the explanations in Swedish rather than in English, and the Es pupils to say whether they would have preferred explanations in English. Many pupils have answered both questions, and they have been scored as "no answer" (Ø) which explains the large number of zero answers here. 26 Im pupils have answered by mistake (or probably many more who have become Ø's as explained above). The Ee pupils are evenly distributed among those who liked English explanations and those who would have preferred to have them in Swedish. In the Es groups there is a strong majority for those who preferred to have them in Swedish, i.e. they answered "no" to the question, they would not have preferred to have them in English. All these answers should be taken with great caution, however.

7 B: No less than 43 Im pupils have answered and they generally feel that the explanations they thought they had got made it much easier for them. It is worth noticing that the Im mean is higher than both those for Ee and Es here! Non-existing explanations thus seem to be easier than real ones! Both Ee and Es have high means(4.17 and 4.16) however, and only 4 and 3 pupils respectively feel that they made it somewhat or much more difficult to understand.

7 C: The Im pupils who have answered feel they had too few explanations. Two of them feel they had a little too many explanations, though! Of the Ee and Es pupils 70 and 106 (52.5 % and 59.4 %) respectively feel that they had just the right amount of explanations. Of the rest most feel that they had too little. Only 11 and 20 respectively feel that there were too many explanations. On the whole this seems to indicate a favourable attitude.

Question 8, finally, which was for the Im pupils only was answered by 22 and 19 Ee and Es pupils. Most of them have answered that they missed explanations sometimes; they are probably the same pupils who answered with a 1 or 2 in 7 C which is quite reasonable. Of the Im pupils 94 (70.5 %) feel that they missed explanations sometimes, 15 never missed them, but only 4 missed them very much. This again indicates a fairly positive attitude to the method without explanations.

To sum up:

From the questions relating to the explanations or the absence of them, it seems that it does not make much difference whether one

Table 36b: Pupils' Attitudes to Explanations.

Question				Frequencies:						
		$\bar{x}$	s	0	1	2	3	4	5	Ø
All pupils	6	.69	.46	163	356					58
N: 577	7A	.67	.47	69	138					370
	7B	4.18	.73		3	4	39	190	118	223
	7C	3.30	.85		12	21	204	84	34	222
	8	2.84	.67		8	30	117	19		403
Im	6	.24	.43	129	41					11
N: 181	7A	.69	.47	8	18					155
	7B	4.26	.54		0	0	2	28	13	138
	7C	3.37	.76		0	2	28	8	5	138
	8	2.89	.63		5	19	94	15		48
Ee	6	.85	.35	23	134					38
N: 195	7A	.50	.50	42	42					111
	7B	4.17	.74		1	3	12	74	43	62
	7C	3.38	.85		4	7	70	40	13	61
	8	2.59	.80		3	4	14	1		173
Es	6	.94	.23	11	181					9
N: 201	7A	.80	.40	19	78					104
	7B	4.16	.77		2	1	25	88	62	23
	7C	3.22	.87		8	12	106	36	16	23
	8	2.79	.71		0	7	9	3		182

6 In my class *we had* explanations (1) - *did not have* explanations. (0)
7A It would have been better with explanations in English/ in Swedish. no=1 yes=0
7B The explanations made it *much more difficult* (1) - *much easier* (5) to understand.
7C He had *too many* (1) - *too few* (5) explanations.
8 *I did not miss* (4) - *I very much missed* (1) explanations.

Note: number 7 was for Ee and Es groups only, number 8 for Im groups only; In 7A Ee groups were supposed to say whether they would have preferred to have the explanations in Swedish, the Es groups whether they would have preferred English.

Note also: Number 7C does not have one positive and one negative end, but rather two negative ends with the more positive answers in the middle.

gives explanations or not. The attitudes of the pupils seem to be nearly the same in both groups. Some who get explanations do not notice them, and some who do not get any think they have got them.

Interest in English.

The pupils' interest in their school subjects was measured by a special interest test. The results of this are discussed in Appendix C since this is slightly outside the general scope of this report. The figures for English will be given here, though.

Interest was measured with a four-graded scale: ++ + - -- , the steps were given numerical values 4 - 3 - 2 - 1, and means calculated. The result for English is as given in table 37.

Table 37 : Interest in English.

Im		Ee		Es	
Class		Class		Class	
1	3.2	10	2.8	19	2.4
2	3.0	11	3.2	20	2.5
3	3.0	12	3.1	21	2.7
4	2.8	13	3.4	22	3.4
5	3.7	14	2.5	23	3.3
6	3.1	15	2.9	24	2.9
7	2.9	16	3.3	25	3.0
8	3.4	17	3.2	26	2.4
9	2.9	18	2.9	27	2.9
Total	3.1	Total	3.0	Total	2.8

Total mean for all classes: 3.0

The total mean of 3.0 for English, which happens to correspond exactly to that for all subjects, shows that on the whole the pupils find school more fun than boring and that this applies to English also. It seems that the Es pupils are the most negative. To what extent this may have influenced the results on the achievement test is impossible to say.

The most positive class, number 5, has a mean of 3.7, which indicates that almost all the pupils find English almost always fun. The other extreme is found in two Es classes with 2.4, which indicates that in these classes most pupils find English rather dull.

These figures indicate most convincingly that in grade 6 English is a popular subject. Curiously enough, in GUME 1, (Lindblad, 1969, p. 84) the extreme values were also 2.4 and 3.7, the means for the easier course in grade 7 being 2.8 and for the advanced course 3.2; the attitude towards English is fairly constant at this age and ratings almost two years apart yield identical results.

Teachers' Attitudes.

General methods questions. The first five questions (numbered 3 - 7) concerned general methodological preferences among the teachers.

Four teachers say that they use Im normally, six use Ee and 17 use Es. Most teachers thus prefer to give explanations and most of them do so in Swedish. The figures show that our feeling that the methods were no more extreme than that they would all find proponents is correct.

How often do they give explanations? Only one does so every lesson, 10 quite often and regularly, 16 sometimes when it is necessary, but no teacher says he or she never gives any explanations. The interesting thing to notice here is that so many seem to explain so seldom. It should be noticed that this is when the pupils are at the end of their third year of English, having had a total of 11 hours a week (2+5+4). In giving explanations 18 prefer to do so themselves rapidly and in a concise way, whereas 9 let some pupil do it first and then round it off themselves. Some say that they use a mixture of these methods. There is a majority, though, for distinct "rules" given by the teacher rather than the inductive generalization.

How much do teachers speak English during their lessons? It is of course difficult to estimate. One teacher says it is 75 %, but "the pupils say it is 90 %": subjective feelings influence them here, but the answers were as follows:

99 %	1
90-95 %	7
80-85 %	10
70-75 %	8
60 %	1
	27

A difference such as between 85 % and 80 % is too small to indicate any real difference but certainly there are differences between the extremes. One teacher obviously does not speak any Swedish at all whereas at least one speaks quite a bit of Swedish in his or her lessons.

How often do teachers in Sweden use structure drills, one of the most characteristic traits of the audio-lingual method?

always	2
quite often and regularly	14
sometimes	10
never	1
	27

Judging from these figures it seems that this method has become accepted and is now widespread at this level. It should be noticed, though, in looking at these figures and at those for some of the other questions, that all teachers in the project used Ashton-Olsson, Hands up, as their textbook. This is a modern book with a very well-written Teachers' Handbook that most teachers follow fairly closely. They get help from it and are influenced by it. It is most likely that classes whose teachers use older textbooks written on more traditional lines would have differed significantly from the project population.

Questions on the project. As in the Pupils' Attitude Test the teachers first answered two open questions in order to bring out spontaneous reactions of what was felt to have been good and bad in the project.

Good. The Im teachers generally felt that it was good for the pupils to have to listen to nothing but English, that they did not have to have "incomprehensible" explanations, that they had structure drills and that there were so many oral drills. The Ee teachers generally liked explanations in English (even those who would normally give them in Swedish themselves), they thought the explanations were good and easy to follow and they liked the oral and written structure drills. The Es teachers liked the explanations and the fact that they were in Swedish.

Bad. Some teachers feel that some grammatical points were not brought home well enough; this seems to be true in particular of 'some-any', and the past tense. Sometimes they feel there were too many repetitions

and too much oral drill. The weaker pupils in Im missed the explanations. Some Ee teachers think explanations in English may have been too difficult. One or two would have liked to have the explanations written out also. One teacher in Es thinks there were too many explanations taking time from the more valuable drills. Almost all the teachers have liked the *oral drills*; some think that sometimes there were too many or too long drills. The written drills are even more fabourably mentioned; three teachers feel that sometimes there was not time enough for the less gifted pupils to finish (which they were not supposed to do). The reading texts are called good or excellent, but many teachers think they were a bit too difficult and that there were new words the pupils did not understand and which therefore irritated them.

Explanations: the Im teachers who missed explanations (not all of them did) mention the s-form and 'some-any' as the points where the pupils needed explanations most. The Ee teachers differ a bit: most seem to have felt that the pupils understood them, but some think it was too difficult and that the pupils were confused. The Es teachers are all satisfied and several mention the simple present and past as opposed to the continuous tenses as a particularly good point.

As to the *tempo* all teachers feel that the speed of speaking was good, and most of them feel that pauses etc were just right. Quite a few found the pauses in the oral drills somewhat too long but at least one thought they were sometimes too short. On the whole the tempo seems to have been all right.

The *technical quality* was good, but some teachers complain that some of the tapes were not first-class, but in no instance has this caused any serious trouble. All materials used have worked well, but in some schools the teachers had trouble finding overhead projectors.

The most common comment about *pupils' interest* is that it was great to begin with but slowly decreased. Only in a few cases did this cause anything like irritation.

What do the teachers think of the *learning effects* prior to knowing the final results? Some do not want to guess but otherwise answers vary very much. "Not very great", "the best pupils probably little", "no doubt quite a bit", "good for the best pupils", "same as usual".

The most interesting thing to notice here is the lack of uniformity in opinions whether the best or the least gifted pupils learnt anything.

(Most teachers were very surprised when they learnt about actual results. One of the most critical teachers, who happened to have a very good class which had made great progress and in which particularly the bright pupils had made great progress, was extremely surprised. This seems to indicate that it is very difficult for a teacher to tell whether the pupils are really learning or not).

The achievement test was generally considered good but very difficult. In commenting on *the lessons* most teachers give very positive answers. They have found them varied and interesting. Some of the criticisms about the lack of explanations etc are repeated. It is felt that especially 'some-any' and the past tense were given too little time. Particularly the last three lessons seem to have been a bit too crammed for some classes.

To end up, the teachers were asked to estimate how they felt time had been used during the project.

	Im	Ee	Es
almost completely wasted	0	0	0
fairly wasted	0	2	1
fairly well used	4	4	4
very well used	4	2	2
no answer		1	2
	8	9	9

One Im teacher felt that the time was completely wasted for the better pupils, fairly well used for the middling and poorer pupils, and very well used as regards herself!

DISCUSSION OF RESULTS

If our earlier investigations (GUME 1-3) are regarded as pilot studies, it may be stated that they are quite comprehensive and meticulous as such. The present study was planned against the background of them; the planning thus gained from the illumination of hindsight. Modifications in design and otherwise were made in order to increase the probability of revealing method differences. It is our subjective judgment that the three teaching methods compared, in GUME 4, the Implicit, the Explicit-English, and the Explicit-Swedish, were altered to the better in comarison with the earlier studies. In spite of this, the main results of the separate studies became the same - the three teaching methods did not generate any differences in learning effects.

The research tradition that the GUME project represents, i.e. method comparisons in an educational setting, is discussed at some length by Stephens (1967). The results generally obtained within this area of research seem to have become a tradition, too (p. 7): "It is part of the folklore that, in educational investigations, one method turns out to be as good as another and that promising innovations produce about as much growth as the procedures they supplant, but no more". To take another example, Nachman and Opochinsky (1958, p. 245) state that "Reviews of teaching research have *consistently* (italics ours) concluded that different teaching procedures produce little or no difference in the amount of knowledge gained by the students". To return to Stephens, he gives a comprehensive survey of method comparisons, the main impression of which is one of negative results (by negative he understands non-significant differences between methods compared). According to Stephens (p. 82 ff), many authors suggest that the negative results should not be accepted as final answers and therefore they point to various reasons for these negative findings. Since some of these might be valid for our results, we shall present them briefly and give our own comments as well:

1. It is pointed out that the experiments test only one narrow segment of achievement, namely those that are easy to test. The argument goes on to say that great changes in other aspects of achievement, especially in personality or character, might be discerned if these were tested.

In the case of GUME 4 the productive oral aspects of linguistic competence are set aside by the Achievement test, and it is of course theoretically possible that a test stressing the oral skill might have given results different from those obtained. However, this outcome is hardly probable considering the correlations between the productive and receptive aspects of language as measured by our tests (see p. 98 above) as well as our technique for grading the productive written tests (see p. 45 above).

2. A second argument contends that the tests used are not only too narrow in their scope, but they are relatively insensitive even in the area in which they do function. This argument implies that more sensitive measures might detect considerable growth which now escapes observation.

   The reliabilites of the various parts of our Achievement test (p. 49) should serve as an acceptable counter-argument to this criticism.

3. In the flood of investigations there is much variation in rigour and scientific care. Many investigations clearly fail to control factors that could have affected the results.

   This source of error is perhaps the most common in research of this kind: vague instructions to participating teachers and pupils, malfunctioning of technical equipment, insufficiently tried-out teaching sequences, changes in experimental schedule because of events which might have been foreseen, variations in listening conditions between classrooms; indeed there are numerous potential causes of irrelevant influence. Without passing value judgments on the present project as far as the execution of the study is concerned, we can say that we were well aware of many obstacles, having one ordeal behind us.

4. A fourth explanation attributes the lack of positive results not to lack of control but to "overcontrol". The educational investigator, in his zeal to become superscientific, has controlled the investigation "to death", so to speak. In his effort to make sure that extraneous factors are held constant, he has held the whole growth process constant.

   It is somewhat difficult to see what Stephens actually means by overcontrol, but if he should be taken to mean an attempt at

working under laboratory conditions rather than in a real-life situation, then GUME 4 is definitely not overcontrolled. The study was carried out in the natural school setting and we tried to change the normal routine as little as possible.

5. In judging whether or not significant positive results exist, a criterion has been used that is much too strict. Often we have refused to admit that a difference is significant unless we can be guaranteed odds of 1 to 100 or 3 to 1000. In the face of a handicap such as this, it is no wonder that many results have been negative. It is a wonder that they have ever been positive.

   Although it is of no consequence what level of significance is applied to our main results, the obtained F-ratios being far from the critical values, the problem at issue is nevertheless important. If the educational researcher entertains a hope that his work shall ever influence school life, he must be prepared to advance plausible, not to say strong arguments for his "cause". For instance, if one of the GUME methods had consistently proved superior, it is very likely that the method had not, because of this, been accepted and proposed as "the method" in the schools. Because the introduction of a new teaching method costs a considerable amount of money (teacher training, production of materials, etc.), it takes strong statistical evidence to get it introduced. The strength of the argument would, among other things, be dependent on the level of significance used in the statistical tests. Thus, considering a hypothetical introduction in the schools of Im, Ee, or Es - whichever proved to be the best - the 1 % level would probably be a necessary prerequisite for convincing the school authorities about the superiority of that method. In a study like the present one the statistical criterion should thus be strict rather than liberal. As it appears, we are in opposition to Stephens here.

What is most interesting about Stephens's critique of the comparative experiment is his contention that differences in the formal method of teaching, compared to the strong influence of and variation in background variables, may have difficulty in demonstrating their influence. "Administrative factors and pedagogical refinements ..... are inevitably left to show their influence on that part of the (learning) curve where diminishing returns are the rule" (p. 85).

We tend to agree with Stephens in his general comment; in the case of the present study it seems more justifiable to accept as a fact that the differences between the methods did not produce corresponding differences in learning effects, rather than invoke the classical hypotheses of imprecision and random error.

Thus, within the total GUME project, evidence has now been accumulated in one and the same direction: method differences such as ours account for very little of the actual variation in learning results. This is, in itself, a most interesting finding against the background of the intense Swedish debate in 1969-70 in which proponents of (mainly) two "schools" defended their respective methods with what might be termed limited tolerance and definitely little support from empirical research.

Stephens's list of reasons for negative results could have been made longer. However, even if we could imagine a completely perfect study from a research point of view, there might still be good chances for negative results to appear: the human brain is obviously a flexible enough structure to allow for learning under a variety of conditions. It is a commonplace that learning occurs even under non-optimal conditions. Fortunately, the learner may understand a message though it is transmitted badly (through the wrong channel, to use the jargon of information theory). Perhaps the reader has noticed occasionally, when confronted with "bad" teaching, that he was able to grasp the message despite the poorly arranged teaching situation. Still in the language of information theory: this ability to decode noisy, even faulty, messages is one indication of the potential and flexibility of the human information processing system. With this in mind one should perhaps not expect modest differences among (hopefully) equally reasonable teaching methods to cause differences in learning effects.

The type of research that GUME represents is thus beset with difficulties; unless one works under laboratory conditions, as did, for instance, Crothers and Suppes (1967), the study seems doomed to negative results, and when one tries to achieve the purity of experiment that they did, one seems to loose all touch with the real-life conditions of second language learning.

It was stated earlier in this report (p. 31) that we hoped our results might shed some light on the debate in methodological matters

which has been, in Sweden as elsewhere, very lively. What kind of illumination can results such as ours yield? It seems that the results indicate that much of the debate, at least as far as it refers to 'grundskolan' (the compulsory, non-streamed comprehensive school), is on the wrong track. The variation in a number of "extraneous" background variables within a given group of pupils and the differences between teachers in personality, training and skills seem to be of such a magnitude that differences between groups taught by different methods are completely levelled out. If this refers to a study such as the present one where the methods used are strictly defined and adhered to, this is in all likelihood much more the case in the ordinary classroom situation where teachers confessing to believe in the same method may very well teach in completely different ways, and vice versa. The quarrel about methodological details thus seems misguided, and attention should be directed elsewhere: the linguistic training of the teachers, the personality of the teacher, the social background of the pupils, reasons for "school-tiredness" in the pupils, the size of classes, technical aids which facilitate individualization, and many other fields like these.

The opinion of some critics that GUME has yielded nothing but "non-results" which is distressing is not shared by the project group. We feel that results that can show the uselessness of one direction of the discussion and thus open up other more fruitful perspectives are indeed valuable and interesting.

THE GUME 1 AND 2 FOLLOW-UP STUDIES 1969/70.

In planning and carrying out the GUME 1 and 2 studies in 1968-1969 all classes taking part were experimental classes. No control groups were used. There were several reasons for this. First of all we used a total of 54 classes in GUME 1-3 which is a large percentage of the classes on this level available in the Gothenburg area (between 20 % and 25 %). Secondly, we were to teach one grammatical structure (GUME 1: the do-construction, 2: some-any, 3: passive) intensively in six lessons and then measure progress. In a class whose teacher did not concentrate on the same structure during this period progress would in all likelihood be close to zero. And if the teacher did concentrate on it, there was no way of checking how he did it and thus what we would be comparing with. For these two reasons mainly, no control groups were used. Furthermore, we did not feel a very strong need for control groups since we were not interested in the amount of raw progress made as such but only in the difference in progress brought about by different methods of teaching irrespective of how great or small this progress was.

Since tests with good reliability were available (.92 and .92 for the GUME 1 and 2 tests respectively) for which we had fairly extensive results to compare with, we felt that it might be of interest to compare our results afterwards with what is normally achieved at the same level *in one whole year*. This is all the more interesting as very little has been done in this field. Some teachers in our projects had said in the questionnaire which they filled out after the projects that they would have done better themselves. But does a teacher know how much his pupils normally learn in, say, two weeks of instruction? Do we ever measure our pupils' progress all that carefully?

There is one study by the UME project in Stockholm trying to establish how much Swedish pupils learn of English grammar in the 7th form, but their results are very uncertain because two different tests were used and different classes were tested. We wanted to test one group of pupils twice with the same test to check the results obtained in the Stockholm study. Before the start of the autumn term, 1969, tests, tapes and instructions were sent to the headmasters of

the same schools that had been used in the original studies. They were asked to distribute them to teachers in their schools who were to teach a group of 7th grade English during the coming year. We specified whether we wanted it to be sk or ak (see p. 61f for an explanation of these terms). We used 6 ak and 12 sk classes which corresponds roughly to the proportion in which pupils choose and it also corresponds to the proportions used in the original studies. We thus tested a representative sample of pupils aged 13. The test was given on the very first lesson in the autumn term. The tests were then collected and marked but the teachers were not informed of the results, and they were not told that we would be coming back at the end of the year.

At the end of May all teachers who had taken part in this follow-up study (some of whom, incidentally, had also taken part in the original project studies) were contacted again and asked to give the test to their pupils about ten days before the end of the school year and without giving any extra teaching in the intervening period (which was only a few days). All GUME 1 follow-up teachers gave the test and returned it to us; in some classes there were new teachers but they gave the tests instead. Tests from 2 ak and 4 sk of the GUME 2 follow-up classes could not be obtained; this means a 30 % drop-out rate but the proportions ak-sk were maintained. There is no reason to believe that the loss was systematic and the results can probably be considered representative.

In processing the results only pupils who have taken both the Pre-test and the Post-test have been included. All test figures that have been compared refer to exactly the same group of pupils.

The over-all results are given in tables 38 and 39, results according to course are given in tables 40 and 41, and some correlations in tables 42,43,44. Discussions of the figures are given in connection with the tables.

Table 38: GUME 1: Follow-up and Original Results (all pupils).

Note: S1 is the sum of parts 1,2, and 3 of the test; S2 the sum of parts 5,6, and 8; S3 the sum of parts 11 and 12. The total is greater than the sum of S1-3 since parts 4,7,9, and 10 are also included.

	FOLLOW-UP (N = 363)						GUME 1 means (N = 330)		
	Pre-test		Post-test		Progress		Pre-test	Post-test	Progress
	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	$\bar{x}$	$\bar{x}$
S1	18.28	5.69	20.77	5.64	2.49	3.98	18.71	20.99	2.27
S2	12.97	5.08	15.36	6.25	2.39	4.06	12.20	15.11	2.91
S3	11.34	3.96	13.08	4.18	1.74	3.25	11.48	12.44	.96
Total	64.33	17.35	73.57	19.85	9.24	10.04	64.08	72.91	8.83

Table 39: GUME 2: Follow-up and Original Results (all pupils).

Note: The test consisted of three parts A, B, and C.

	FOLLOW-UP (N = 220)						GUME 2 means (N = 317)		
	Pre-test		Post-test		Progress		Pre-test	Post-test	Progress
	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	$\bar{x}$	$\bar{x}$
A	14.36	6.80	20.01	8.16	5.65	5.78	17.15	23.28	6.13
B	9.70	4.66	12.60	5.10	2.90	3.53	11.33	14.00	2.67
C	28.74	9.61	36.77	10.58	8.04	10.03	31.87	38.25	6.38
Total	52.80	17.62	69.38	20.76	16.58	14.72	60.35	75.53	15.18

Comments on Tables 38 and 39. The GUME 1 test had a maximum score of 120. The Pre-test results in the project - which started about four weeks after the beginning of the term - are equal to those in the follow-up study. The do-construction has been dealt with in grades 5 and 6 and very little obviously happens in the first few weeks. The GUME 2 test had a maximum score of 131 and here we notice that the Pre-test figures in the project, which did not start until November, are 10-15 % higher than in the follow-up study. The some-any problem has not been dealt with systematically before grade 7, and here the pupils progress markedly in the first two months.

The Progress score in GUME 1 is almost identical with that in the follow-up study, which means that the pupils learnt as much in the six project lessons as they do otherwise in one year concerning this particular but important grammatical structure. The progress is about 12 %. This figure is substantial and thus shows that the pupils make real progress in this area, which is in opposition to the results in the Stockholm study referred to above.

The GUME 2 Progress score is also the same for the project and for the follow-up classes, but then it should be remembered that the pre-test figures were different. This means that the post-test figures in the project are also higher than in the control group. Progress is here about 20 %. This greater progress as compared to that for GUME 1 is probably due to the fact that the pupils knew less at the start. The means on the GUME 1 and 2 pre-tests are fairly close, however, (64.08 and 60.35); the pupils have thus answered a little more than 50 % of the questions correctly. Since the pupils knew less of 'some-any' than of the do-construction, this might seem strange. The reason is that the GUME 1 test is more difficult, of course, and also that the GUME 2 test is all of the multiple-choice kind whereas some parts of the GUME 1 test are more active.

To sum up:

The over-all picture is that pupils in grade 7 progress in their knowledge of English grammar, about 10-15 % in the case of the do-construction, which is a central but difficult structure, practised before, about 20 % in the case of the 'some-any' dichotomy, which is an easier and also almost completely new phenomenon. The difference in progress between individual pupils is very great as the standard deviations indicate. In the GUME 1 follow-up group it even exceeds the mean. This means that many pupils make regress rather than progress, and also that some pupils make progress much greater than most of their friends. We shall see below how ak and sk vary in this respect.

Table 40: GUME 1: Follow-up and Original Results for ak and sk.

	FOLLOW-UP				GUME 1 means	
	ak (N = 93)		sk (N = 270)		ak (N=100)	sk (N=225)
	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	$\bar{x}$
Pre-test	48.66	10.39	69.73	15.94	48.95	70.59
Post-test	55.65	12.05	79.74	18.20	52.76	81.86
Progress:						
S1	2.54	3.94	2.47	3.99	2.11	2.40
S2	1.82	3.22	2.59	4.29	.51	5.44
S3	1.57	3.15	1.80	3.29	- .07	1.44
Total	6.99	8.33	10.02	10.47	4.33	11.27

Table 41: GUME 2: Follow-up and Original Results for ak and sk.

	FOLLOW-UP				GUME 2 means	
	ak (N = 66)		sk (N = 154)		ak (N=86)	sk (N=230)
	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	$\bar{x}$
Pre-test	43.79	11.53	56.66	18.38	47.18	65.26
Post-test	53.35	15.75	76.25	18.78	59.31	81.53
Progress:						
A	2.47	5.67	7.01	5.29	4.73	6.70
B	1.56	3.46	3.47	3.41	2.00	2.93
C	5.53	12.21	9.11	8.77	5.27	6.80
Total	9.56	16.89	19.59	12.59	12.00	16.43

Comments on Tables 40 and 41. As in tables 38 and 39 we notice that the Pre-test figures for GUME 1 are identical for the project population and that of the follow-up study and for GUME 2 higher in the project group. We also see that the difference in GUME 2 between project and follow-up results is greater in sk than in ak (8.60 points as compared to 3.39). When we come to Post-test and Progress figures the two projects give contrasting pictures. In GUME 1 the ak group made less Progress than the control group (4.33 as compared to 6.99) whereas in sk project classes did better (11.27 and 10.02 respectively). This is probably due to the fact that the teaching material used in the project was the same for both courses and it was obviously too difficult for the less gifted children. In the ordinary classes special simplified textbooks are used in ak.

In GUME 2, however, the opposite picture is given. In ak the project classes score higher than the control groups (12.00 and 9.56) whereas in sk the opposite is true (16.27 versus 19.59). The reason here could be either the opposite of that proposed for GUME 1, namely that the project materials produced in the project was on the easy side. Another possible explanation is that in sk the pupils have picked up quite a bit of the new stuff in the first few weeks that had passed before the project started. In sk the project classes are 8.60 above the control classes on the Pre-test, and if these points, which represent what was learnt in the first quarter of the school year, are added to their Progress, they exceed the control group quite considerably.

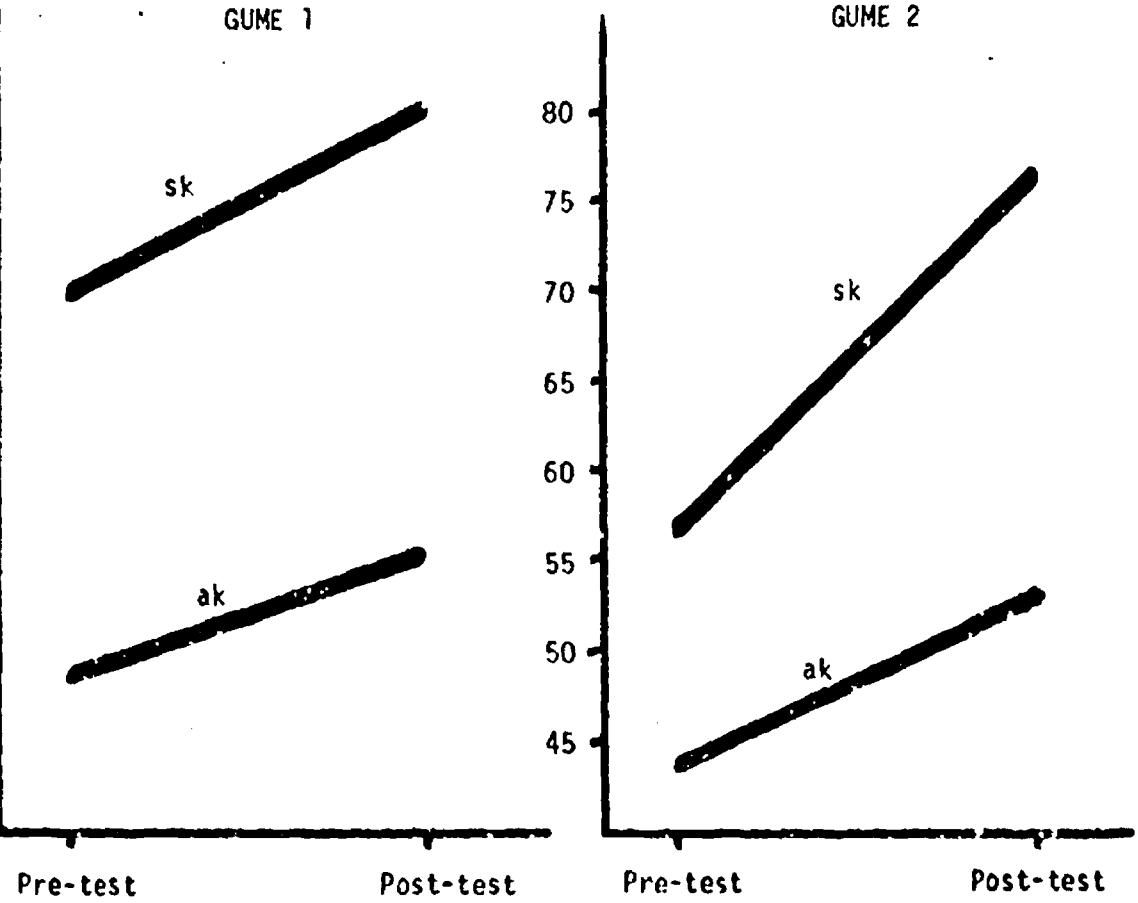
To sum up:
The figures discussed here are in line with the well-established fact that more intelligent pupils not only know more but also make more progress than less gifted ones, thus increasing the difference between two groups of this kind. The comparison between the two projects also points out the importance of producing teaching materials which are easy enough for the poorer pupils. They may also indicate that in the case of a structure which, like the do-construction, the pupils have worked with quite considerably before, the less gifted pupils have reached their ceiling and the learning curve is already bending towards the horizontal or even, for many pupils, downwards, whereas in the case of a new problem it is still showing a strong upwards bend (see fig. 12).

It should also be noticed that in ak the standard deviation is greater than the mean, which is particularly true of GUME 2, both for the project and the follow-up pupils. There is thus a fairly large number of pupils who regress rather than progress both in a short project and over the whole school year.

Frequency Distribution.

In GUME 1 and 2 we found a large overlap between the two courses (ak and sk) both in intelligence and in knowledge of English (Lindblad, 1969,pp.44, 60, Carlsson, 1969,pp. 5, 21, and Levin, 1969, p. 68 f). The latter was also much more pronounced than the overlap in grades, which was taken to indicate that the choice of course, to a large

Fig. 12 : Progress in One School Year (sk and ak Follow-up Pupils)



Comments on the figure:

As the above figure shows, progress in the case of the do-construction is small but marked and about equal in ak and sk. In the case of the almost completely new 'some-any' problem ak progresses at a speed almost equal to that on the do-construction (which is an old, "well-known" structure) but sk differs significantly in that its learning curve rises very sharply. It should be stressed again that this does not represent the somewhat unnatural conditions of an experiment but what pupils learn in the 7th form under ordinary conditions.

extent, is explained by more or less irrelevant factors like feeling of success and social class. See also the correlation figures and discussion of these, pp. 93-102.

Figures 13 and 14 show the distribution of raw scores in the GUME 1 and 2 follow-up studies for the two courses on the Pre- and Post-tests respectively. The lowest figures in GUME 1 in the Pre-test are 28 and 31 for ak and sk respectively, the highest 72 and 117. Lowest on the Post-test are 30 and 38, highest 86 and 118. The spread is thus great, greatest in sk. The difference on the two tests fairly moderate; compare the results section p. 120.

In GUME 2 the lowest figures on the Pre-test are 16 and 23 for ak and sk respectively, the highest 73 and 112. On the Post-test the lowest are 25 and 36, the highest 87 and 119. The spread here is slightly greater than that in GUME 1, and about 30 % greater in sk than in ak. The difference on the Pre- and Post-tests, as opposed to GUME 1, is fairly large, especially in sk.

A problem which we discussed in the previous reports was to what extent pupils choose the wrong course. There are many different criteria to decide this. The most conservative but also most realistic seems to be Anastasi's (1958, p. 454) that only those who fall beyond the median of the other group are in the wrong group. These are shaded in the figure. The figures for GUME 1 then show that very few ak pupils (as a matter of fact fewer than in last year's study) exceed the sk mean. There are 2 in the pre-test and 3 in the post-test. But on the other hand there seems to be many pupils who, in spite of poor knowledge of English, have chosen the more difficult sk. They are 21 and 23 in the two tests respectively.

The number of pupils in GUME 2 who have chosen the wrong course is much greater than in GUME 1. There are relatively few ak pupils beyond the sk median: 7 and 4 for the pre- and post-tests respectively, i.e. 10.8 % and 6.0 %. But there are many sk pupils who score low, especially on the pre-test: 45 and 16 on the two tests respectively, i.e. 29 % and 10.3 %.

We have said elsewhere that it is common statistical experience that more intelligent pupils make greater progress than poorer pupils. This tends to increase the difference between selected groups. In our two groups, ak and sk, we were thus expecting to find a "gliding-apart"

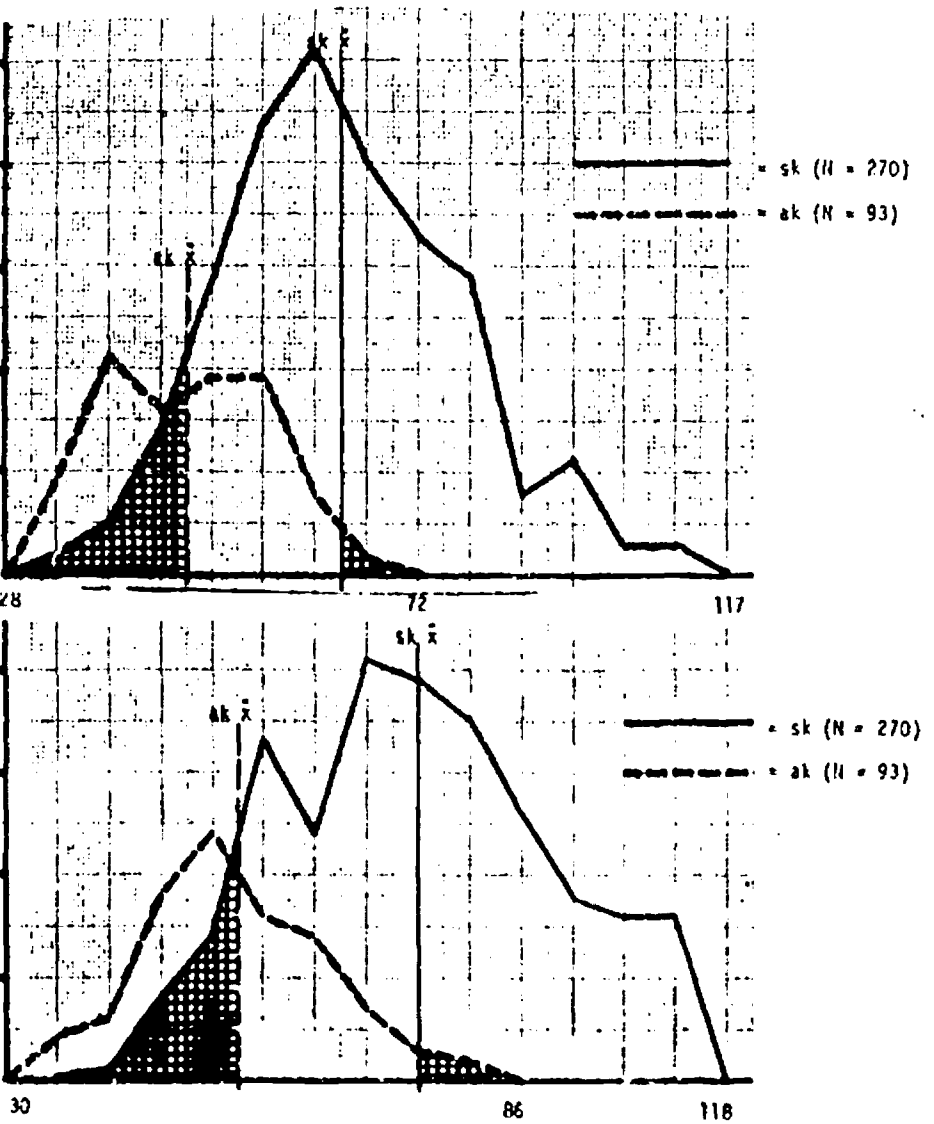


Fig. 13 a.& b: The GUME I Follow-up = Distribution of Scores on
a) the Pre-test (August, 1969)
b) the Post-test (May-June, 1970)

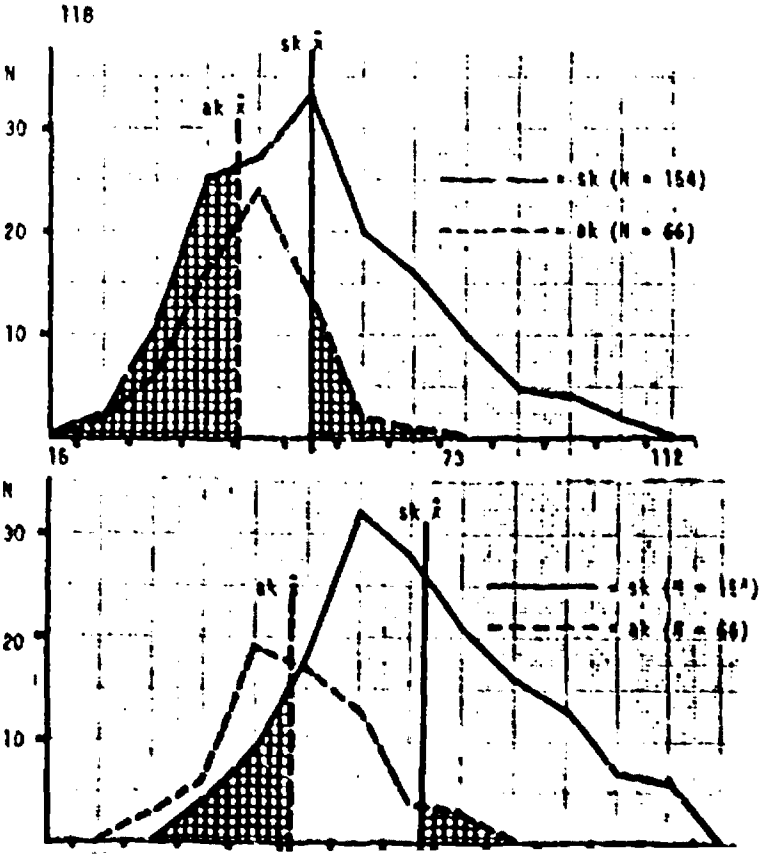


Fig. 14 a & b: The GUME II Follow-up = Distribution of Scores on
a) the Pre-test (August, 1969)
b) the Post-test (May-June, 1970)

effect in the frequency distribution figures. In the case of GUME 1 this effect is hardly noticeable: we can see that the sk mean has moved towards the right a little more than that for ak, but otherwise the two figures are almost identical. In GUME 2, however, the expected effect is easy to see: not only has the means moved apart, but the large overlap has diminished considerably. The fact that so many sk pupils did not know about 'some-any' at the beginning of the year is hardly surprising, considering that they had never learnt about this systematically before. After one year of teaching the sk pupils have moved ahead, though.

It is reasonable to expect that this "gliding-apart" effect becomes more marked as the years in 'högstadiet' go by.

It is obvious that if all pupils who, according to the criterion used above to decide which pupils have made the wrong choice, should change, we would get more pupils in ak than we have at present. If, on the other hand, we draw the line simply where the two curves intersect, and say that those who are to the right of that line should be in sk and vice versa, then we would get very few pupils in ak.

One conclusion that seems valid with the above figures in mind is that the pupils' choice of course should perhaps be guided a little more actively than seems to be the case at the moment.

Some Correlations.

A number of correlations have been calculated to compare with last year's figures. As tables 42 and 43 show (table 42 is partly incomplete since not all figures were obtained in the previous study) correlations with the various parts of the tests are almost identical when the experimental population and the follow-up groups are compared. It can be noticed for example that in GUME 1 the group of three tests called S2 and in GUME 2 test 8 in the Pre-tests correlate with the Post-test totals .787 and .777 respectively, which means that about 60 % of the final variance is predicted by these small tests which take something like 7 or 8 minutes to administer. The reliability coefficients for these two parts of the two tests are .81 and .76 respectively which is quite satisfactory for group comparisons. The reliability coefficients of the whole tests are .90 for both (as compared to .92 for the experimental groups) which is good enough for prognostic and diagnostic purposes with individual pupils.

Table 42 : GUME 1: Test Correlations (all pupils).

		Follow-up							GUME 1		
		Pre-test			Post-test				Pre-test		
		S2	S3	Total	S1	S2	S3	Total	S2	S3	Total
Pre-test	S1	.735	.682	.898	.754	.663	.644	.765	.727	.756	.900
	S2		.698	.891	.732	.762	.626	.787		.691	.879
	S3			.842	.709	.663	.682	.756			.877
	Total				.810	.779	.728	.863			
Post-test	S1					.765	.708	.897			
	S2						.687	.910			
	S3							.846			

Table 43 : GUME 2: Test Correlations (all pupils).

		Follow-up							GUME 2						
		Pre-test			Post-test				Pre-test			Post-test			
		B	C	Total	A	B	C	Total	B	C	Total	A	B	C	Total
Pre-test	A	.766	.517	.870	.716	.653	.617	.756	.734	.485	.856	.765	.692	.607	.783
	B		.382	.768	.764	.743	.578	.777		.380	.755	.775	.741	.518	.752
	C			.846	.250	.185	.510	.403			.844	.352	.351	.680	.578
	Total				.615	.549	.669	.717				.697	.657	.748	.822
Post-test	A					.820	.603	.902					.827	.569	.888
	B						.501	.823						.508	.828
	C							.870							.866

In the experiment no progress correlations were calculated but in the follow-up study they were as follows in table 44 (see next page).

As the figures in this table show there is no correlation between results on the Pre-test and Progress which means that those who scored high have made the same (but not better) Progress than those who scored low. In some instances, e.g. GUME 2, part C, the correlation is even negative. These figures differ from those presented in this report on GUME 4, where there are small but significant positive correlations. Whether this difference has come about through sheer chance or whether the fact that GUME 4 represented an intensive teaching and

Table 44: GUME 1 and 2 Follow-up Pupils: Progress Correlations (all pupils).

		GUME 1 Follow-up Progress				GUME 2 Follow-up Progress			
		S1	S2	S3	Total	A	B	C	Total
Pre-tests	S1/A	-.36	.10	-.00	-.04	-.17	-.07	.16	.03
	S2/B	-.01	-.08	-.05	.02	.18	-.25	.24	.18
	S3/C	.03	.15	-.34	.04	-.26	-.24	-.42	-.44
	Total	-.14	.08	-.09	-.02	-.16	-.22	-.11	-.19
Post-tests	S1/A	.34	.26	.05	.37	.57	.18	.40	.54
	S2/B	.14	.59	.08	.45	.40	.47	.35	.50
	S3/C	.08	.27	.46	.41	.13	-.04	.57	.43
	Total	.18	.42	.17	.49	.38	.16	.53	.55
Progress	S1/A		.20	.07	.59		.33	.38	.73
	S2/B			.17	.68			.19	.50
	S3/C				.49				.87

learning period of a month and this follow-up study a more normal, slower growth of knowledge, we have no way of telling at the moment. The difference is interesting, however, and well worth a closer checking in the future.

Post-test correlations with Progress are high and significant, of course, but this is hardly surprising.

The progress on the various parts of the tests correlate with different magnitude with the total progress. In GUME 1 the figures vary .19 (.49 and .68 being the lowest and the hignest). In GUME 2 they vary .37, which is a marked difference. Test C alone explains about 75 % of the variance if that tests were used alone.

To sum up:
The general impression of the correlations is the same as for GUME 4: they are high (except in the case of some figures in the Progress matrix as discussed above) and even. The high Pre-test - Post-test correlations indicate that the test has high reliability. All parts correlate well with each other; they thus measure the same thing, knowledge of a certain aspect of English. The internal validity is good.

SUMMARY

The present investigation is a direct continuation of earlier GUME studies. Since these produced non-significant differences between three teaching methods compared, it was considered worthwhile to perform a new experiment with a modified design and with any other kind of modification that might increase the probability of detecting true differences between methods, if such existed.

The teaching phase of the present study, abbreviated GUME 4, took place in April, 1970, and consisted of a series of twelve lessons in which various grammatical structures in English were taught. The pupils were in their third year of English (grade 6, approximately 13 years of age).

The independent variables of the experiment were three teaching methods, namely

Im The Implicit method

Ee The Explicit-English method

Es The Explicit-Swedish method

Although the names of the teaching strategies are the same as in the previous studies (GUME 1-3) the teaching procedures were altered to some extent. Thus, in the case of GUME 4 the time for explanations varied between Ee and Es. A strong need was felt for the E methods to contain "optimal" explanations even if this meant a certain variation in explanation time, causing some looseness in experimental control. *The Implicit method* corresponds to an audio-lingual method without generalizations, *the Explicit-English method* corresponds to an audio-lingual method with direct-method generalizations in the target language, *the Explicit-Swedish method* corresponds to an audio-lingual method with explanations or generalizations in the source language; comparisons with corresponding structures in Swedish were also made.

In the study 27 school classes took part, 9 per teaching strategy. Data were processed for a total of 577 pupils. The school classes were randomly assigned to teaching method, the only restriction on the procedure being that no two classes within the same school were allowed to get the same method.

Three parallel lesson series (Im/Ee/Es) were constructed, each consisting of 12 lessons. In order to control the teacher factor, "canned" lessons were used throughout the experiment. However, the teachers were instructed to inspire the pupils, in a strictly prescribed way, to take an active part in the work, especially in the case of oral drills; this was done by way of pointing, gestures, etc. In each classroom extra loudspeakers were installed to improve listening conditions.

In rough outline the experimental schedule was as follows: IQ test, distribution of materials to the schools, Pre-test, the lesson series (i.e. the experiment proper), Post-test, Pupil and Teacher Attitude tests, Standardized Test in English, PACT (a listening comprehension test), conference with the participating teachers.

Progress during the experiment was measured as the difference between the Post-test and the Pre-test scores. The Achievement test was constructed so as to correspond to the particular objectives of the present investigation. It covered the various grammatical structures taught and contained 160 items in all.

The IQ test was the so-called DBA-test (Differentiell Begåvnings-Analys = Differential Intelligence Analysis). The reason for administering this test of general intelligence, was partly to use it as a background variable in some of the analyses and partly to divide the pupil population into three levels of ability and investigate interaction between teaching method and intelligence level.

In the statistical treatment of the data only pupils who were present 10-12 lessons were included; this is equal to stating that those who were absent from three lessons or more were not included in the calculations. Various checks on the drop-outs thus defined (absent three lessons or more) showed that they did not deviate from the experimental population in background variables; thus there is reason to belive that absence was due to chance (illness, visits to the school dentist, and the like). The only statistically significant difference found between the experimental population and the drop-outs was in Progress where the experimental group scored highest. This is taken as a clear indication that the instructional program worked well - it paid to be present during the lessons.

The standing of the experimental group on some relevant background variables (IQ, Grades, the Standardized Test in English) was checked. The group is near the norm on most measures and is therefore considered

sufficiently representative of pupils in grade 6 for the results to be generalizable to that population.

The total progress in raw scores during the experiment was 17.26 points; there is thus ample room for teaching method differences, if any, to appear.

In a number of analyses of covariance Progress was the dependent variable and various background measures (IQ, Pre-test, the Standardized Test, PACT) were used as covariates. Similar analyses were performed separately at the Upper and Lower levels of intellectual ability. Likewise, an analysis of covariance was performed with the Post-test as the dependent variable and the Pre-test as the covariate. In all these analyses the three teaching methods, Im/Ee/Es, proved to be equally effective; the F-ratios were generally so low as to make consideration of tendencies among the absolute figures meaningless.

Two analyses of variance (two-way) were performed in order to investigate interaction between teaching method and ability level (in one case Progress was the dependent variable, in the other the Post-test). No interaction was found.

The analyses mentioned thus far were made with individual scores as the unit of analysis. Some complementary analyses were performed with the school class mean as the unit of analysis. However, these calculations strengthened the impression of non-significant differences between the treatments. Differences did exist, though between school classes within methods.

Two additional measures of Progress were calculated, both relating the pupil's Progress score to his score on the Pre-test. However, these types of scores did not give any results deviating from those obtained for raw scores.

A more detailed analysis was made of the different parts of the Achievement test; certain items in each part test were chosen for further scrutiny. These items, called "critical items", were felt to maximize method differences (for instance, the "critical items" of different part tests might vary in progress for different methods). However, the general picture of equality between the methods applies also at the part-test level.

The pupils' attitudes to the project leaned towards the positive. Certain parts of the questionnaire have obviously given non-reliable

information; this is particularly true of the questions on the explanations used in the lessons. Some pupils who were given explanations did not notice them, and some who did not get any thought they had had explanations.

Thus the main results of the present study are entirely in line with those obtained in our earlier investigations. It seems to make surprisingly little difference which of the three teaching methods is used.

- - - - - - - - - - - - - - - -

Independently of the present study a number of control classes were studied to find out how much of the contents of a GUME course is learnt during one school year without the teacher's paying concentrated attention to the particular grammatical structures (as was done in the experiment). The GUME 1 and GUME 2 courses, i.e. the do-construction and some/any respectively, were chosen for this comparison. In both these cases the original experiments lasted 6 lessons. As it appeared, the pupils learnt as much in the six project lessons as they do otherwise in one year concerning a particular but important grammatical structure.

BIBLIOGRAPHY

Earlier GUME Reports.

Lindblad Torsten (1969) *Implicit and Explicit. An Experiment in Applied Psycholinguistics, Assessing Different Methods of Teaching Grammatical Structures in English as a Foreign Language.* GUME-projektet 1. Lärarhögskolan och Universitetet i Göteborg.

Carlsson Ingvar (1969) *Implicit and Explicit. An Experiment...* GUME-projektet 2.

Olsson Margareta (1969) *Implicit and Explicit. An Experiment...* GUME-projektet 3.

Levin Lennart (1969) *Implicit and Explicit. A Synopsis of Three Parallel Experiments in Applied Psycholinguistics, Assessing Different Methods of Teaching Grammatical Structures in English as a Foreign Language.* The GUME Project. Department of Educational Research, Gothenburg School of Education.

Other References.

Aleamoni L M & Spencer R E (1969) "An Evaluation of the Pennsylvania Foreign Language Project". *The Modern Language Journal.* Vol. LIII, No. 6 (October, 1969).

Anastasi A (1958) *Differential Psychology.* New York: Macmillan.

Anderson R C (1969) "The Comparative Field Experiment: An Illustration from High School Biology". *Proceedings of the 1968 Invitational Conference on Testing Problems,* Educational Testing Service.

Andrakammarvalet i Göteborg 1968. Stickprovsundersökning. (A Sample Study of the Gothenburg Electorate in The Parliamentary Election in 1968) Göteborgs stads statistiska kontor. U 1969:2.

Anvisningar till standardproven för grundskolans årskurs 6. (1969) Stockholm: Skolöverstyrelsen. (Manual for the Standardized Tests in Grade 6)

Ashton J & Olsson Y (1965) *Hands up! 6. Textbok.* Lund: Gleerups.

Ashton J & Olsson Y (1965) *Övningsbok till Hands up!* Lund: Gleerups.

Ashton J & Olsson Y (1965) *Lärarhandledning till Hands up!* Lund: Gleerups.

Brooks, Nelson (1960) *Language and Language Learning.* New York: Harcourt, Brace & World.

Campbell D T & Stanley J C (1963) "Experimental and Quasi-experimental Designs for Research on Teaching". In *Handbook of Research on Teaching,* ed.Gage (1963), pp. 171-246.

Carroll J B (1958) "A Factor Analysis of Two Foreign Language Aptitude Batteries". *J. gen. Psychol.,* 59 (1958), pp. 3-19.

Carroll J B (1963) "Research on Teaching Foreign Languages". In *Handbook of Research on Teaching,* ed. Gage (1963), pp. 1060-1100.

Carroll J B (1965) "The Contributions of Psychological Theory and Educational Research to the Teaching of Foreign Languages". *MLJ,* Vol. XLIX, No. 5 (September, 1965).

Carroll J B (1966) "The Contributions of Psychological Theory and Educational Research to the Teaching of Foreign Languages". In *Trends in Language Teaching,* ed. Valdman (1966). (Revised version of Carroll, 1965)

Carroll J B (1969) "What Does the Pennsylvania Foreign Language Research Project Tell Us?" *Foreign Language Annals,* Vol. 3, No. 2 (December, 1969).

Casey D J (1968) *The Effectiveness of Two Methods of Teaching English as a Foreign Language in Some Finnish Secondary Schools.* Research Bulletin No. 24, Institute of Education, University of Helsinki.

Clark J L D (1969) "The Pennsylvania Project and the 'Audio-Lingual vs. Traditional' Question". *MLJ,* Vol. LIII, No. 6 (October, 1969).

Cronbach L (1963) "Evaluation for Course Improvement". *Teachers' College Record,* 64, No. 8 (May, 1963).

Crothers E & Suppes P (1967) *Experiments in Second-Language Learning.* New York: Academic Press.

Edwardsson R (1970) *Språkdebatten 1969 - 1970*. Stockholm: Utbildningsförlaget. (The Language Teaching Debate, 1969-1970)

Ellegård A (1970 a) "Dåliga språkkunskaper är direktmetodens fel". *Dagens Nyheter*, June 12, 1970.

Ellegård A (1970 b) "Nej, men strukturövningarna!" *DN*, July 30, 1970.

Gage N L (ed) (1963) *Handbook of Research on Teaching*. Chicago: Rand McNally & Co.

Grittner F M (1968) "Letter to the Editor". *Newsletter of the National Association of Language Laboratory Directors*, Vol. III, No. 2 (December, 1968).

Hensjö P-O (1970 a) "Varning för glosplugg". *DN*, June 24, 1970.

Hensjö P-O (1970 b) "Ska grammatiken kastas ut?" *DN*, July 30, 1970.

Hocking E (1969) "The Laboratory in Perspective: Teachers, Strategies, Outcomes". *MLJ*, Vol. LIII, No. 6 (October, 1969).

Härnqvist K (1969 a) *Nya normer för DBA årskurs 6*. Stockholm: Skandinaviska Testförlaget. (New Norms for DBA, grade 6)

Härnqvist K (1969 b) *Resultat från undersökningar med DBA och DBA-liknande test*. Stockholm: Skandinaviska Testförlaget. (Results from Investigations with DBA and Similar Tests)

Larsson L & Sandgren B (1968) *En studie av kreativitetsutvecklingen inom årskurserna 4-9 samt en undersökning av kreativitetens samvariation med intelligens*. Lic.-avh. University of Gothenburg. (A Study of the Development of Creativity within Grades 4-9 and an Investigation of the Covariation of Creativity with Intelligence). Unpublished.

Läroplan för grundskolan (1962): *"Lgr 62"* (Authorized Curriculum for Swedish Schools) Stockholm: Skolöverstyrelsen.

Läroplan för grundskolan (1969): *"Lgr 69"* (ditto) Stockholm: Skolöverstyrelsen.

Läroplan för grundskolan, *Lgr 69, Supplement Engelska* (1969). (ditto, Supplementary Comments on English as a Foreign Language) Stockholm: Skolöverstyrelsen.

Mackey W F (1965) *Language Teaching Analysis*. London: Longmans.

Nachman M & Opochinsky S (1958) "The Effects of Different Teaching Methods: a Methodological Study". *J. educ. Psychol.*, 49 (1958), pp. 245-249.

Otto F (1969) "The Teacher in the Pennsylvania Project". *MLJ*, Vol. LIII, No. 6 (October, 1969).

Palmer H E (1964) *The Principles of Language-Study*. London: Oxford University Press, 1964. A reprint of the 1922 text.

Rivers Wilga M (1964) *The Psychologist and the Foreign-Language Teacher*. Chicago: The University of Chicago Press.

Rivers Wilga M (1968) *Teaching Foreign-Language Skills*. Chicago: The University of Chicago Press.

Scriven M (1968) *The Methodology of Evaluation*. AERA Monograph Series of Curriculum Evaluation, No. 1, Chicago, 1968.

Smith P D & Berger E (1968) *An Assessment of Three Foreign Language Teaching Strategies*. Washington: US Department of Health, Education & Welfare.

Smith P D & Baranyi H A (1968) *A Comparison Study of the Effectiveness of the Traditional and Audiolingual Approaches to Foreign Language Instruction Utilizing Laboratory Equipment*, Washington: US Department of Health, Education & Welfare.

Smith P D (1969 a) *A Comparison Study of the Effectiveness of the Traditional and Audiolingual Approaches to Foreign Language Instruction Utilizing Laboratory Equipment*. Washington: US Department of Health, Education & Welfare.

Smith P D (1969 b) "The Pennsylvania Foreign Language Research Project: Teacher Proficiency and Class Achievement in Two Modern Languages". *Foreign Language Annals*, Vol. 3, No. 2 (December, 1969).

Smith P D (1969 c) *In Reply to the October, 1969, Modern Language Journal*. A talk by Dr. Philip D. Smith, Jr., 3rd Annual Meeting of the American Council on the Teaching of Foreign Languages, New Orleans, November 28, 1969. Mimeographed manuscript, a revised version of which is included in Smith, 1969 a.

Stanley J C (1970) "Controlled Experimentation: Why Seldom Used in Evaluation?" In *Proceedings of the 1969 Invitational Conference on Testing Problems*, Educational Testing Service, pp. 104-108.

Stephens J M (1967) *The Process of Schooling*. New York: Holt, Rinehart and Winston.

Titone R (1968) *Teaching Foreign Languages. An Historical Sketch*. Washington: Georgetown University Press.

Travers R M W (1960) *An Introduction to Educational Research*. London: Macmillan.

En undersökning av elevernas behärskning av grammatiska färdigheter i årskurs 7. Rapport från "The Pilot Study" i oktober 1966. (An Investigation of the Pupils' Command of Grammatical Skills in Grade 7) Projektet UME, Lärarhögskolan i Stockholm. Mars 1967. Unpublished.

En undersökning av elevernas behärskning av grammatiska färdigheter i årskurs 7. II. En redovisning av uppföljningsomgången i maj 1967. (An Investigation of the Pupils' Command of Grammatical Skills in Grade 7, Part II) Projektet UME, Lärarhögskolan i Stockholm. Oktober 1967. Unpublished.

Valdman A (ed) (1966) *Trends in Language Teaching*. New York: McGraw Hill.

Valette Rebecca M (1967) *Modern Language Testing*. New York: Harcourt, Brace & World.

Valette Rebecca M (1969) "The Pennsylvania Project. Its Conclusions and Its Implications". *MLJ*, Vol. LIII, No. 6 (October, 1969).

Wiley D E (1969) "A Methodological Review of the Pennsylvania Foreign Language Research Project". *Foreign Language Annals*, Vol. 3, No. 2 (December, 1969).

Wittrock M C & Wiley D E (ed) (1970) *The Evaluation of Instruction*. New York: Holt, Rinehart & Winston. Forthcoming.

List of Appendices:

Appendix A: The Achievement Test given as pre- and post-tests. The whole test was recorded and took 2 lesson periods of 40 minutes to administer.

Appendix B: The Pupil Attitude Test administered at the end of the project immediately after the post-test.

Appendix C: Short description of and comment on the outcome of the pupil attitude test concerning interest in all compulsory school subjects.

Appendix D: The Teacher Attitude Test, consisting of two parts, one concerning general methodology, the other the project, answered at the time of the pupil attitude test.

Appendix E: A list of participating teachers in alphabetical order.

Appendix F: Descriptive statistics per school class (N = 27) and for the total experimental population (N = 577).

Appendix A

THE ACHIEVEMENT TEST

GUME-projektet
Torsten Lindblad - Ingvar Carlsson

Prov i engelska

Namn: _____

Klass: _____ Skola: _____

Lärare: _____

Datum: ___/___19___

Lärarhögskolan i Göteborg
mars 1970

DELPROV A

I meningarna till höger här nedanför fattas hela vägen ett ord. Det skall du fylla i. Tag det understrukna ordet i meningen till vänster men ändra formen på det när så behövs. Här är ett exempel:

Do you like music? — *Yes, I like music very much.*

1. Does Ann like dolls?
   Yes, but she _____ *cars better.*
2. What did Mary laugh at last night?
   She _____ *at the film.*
3. Does your father live in Oslo?
   No, he _____ *in Gothenburg.*
4. Does Mack wash his face?
   Yes, he _____ *his face every day.*
5. When did the letter from Ann arrive?
   It _____ *yesterday.*
6. What did you play this morning?
   I _____ *football this morning.*
7. What programmes does Sam watch on TV?
   He _____ *cowboy films.*
8. Where does Kate want to go?
   She _____ *to go to Africa.*
9. Did she want to go by car?
   No, she _____ *to go by plane.*
10. What does Ann do in the mornings?
    She _____ *her homework.*

VÄND INTE BLAD FÖRRÄN DU BLIR TILLSAGD !

GUME Ld-IC - 3/70.

DELPROV B

Tänk dig nu, att du talar direkt till mig som gjort det här provet, och ställ frågor till mig på engelska. Om jag på engelska säger Ask me if I am ill, så bör du ställa frågan *Are you ill?* Gör nu likadant här!

1. Ask me if I walk to school.

to school?

2. Ask me if Bill posted the letter.

the letter?

3. Ask me if Peter plays the guitar.

the guitar?

4. Ask me what Bill and Kate shouted to the dog.

What *to the dog?*

5. Ask me if Tom sings well.

well?

6. Ask me if Susan watches TV every evening.

TV every evening?

7. Ask me when his brother arrived.

When *?*

8. Ask me what Tom does on Sundays.

What *on Sundays?*

9. Ask me why John carries an umbrella on the beach.

Why *an umbrella on the beach?*

10. Ask me where the old man lives.

Where *?*

11. Ask me if the policeman talked to the thief.

to the thief?

12. Ask me if my brother plays the piano.

the piano?

13. Ask me if he ever looks at all his stamps.

at all his stamps?

14. Ask me if I like to listen to pop music.

to listen to pop music?

15. Ask me if Kate goes to school by train.

to school by train?

GUME Ld-IC - 3/70.

DELPROV C

I det här delprovet ingår flera uppgifter. I meningarna finns en hel del luckor. I början av varje uppgift står de fyra ord som du i den uppgiften har att välja emellan. På raderna i meningarna skall du bara skriva den bokstav (a, b, c eller d) som står ovanför det ord som du tycker passar in; du skall alltså inte försöka skriva in ordet självt, för det får inte plats. Här är ett exempel:

a.	*b.*	*c.*	*d.*
am	*are*	*is*	*was*

Mary __c__ a girl. Peter and John __b__ her brothers. I __a__ their father. Last summer Mary __d__ in America. Her uncle __c__ a cowboy there.

Uppgift 1: Här är de fyra orden att välja bland för den här uppgiften:

a.	*b.*	*c.*	*d.*
dance	*dances*	*dancing*	*is dancing*

Peter is not very fond of _____, but he is fond of Mary, and she _____ so well that Peter can _____ for hours when he _____ with her. In this picture he _____ with her at a party in Pat's house. (Nästa dag talar Mary med Betty.)

Mary: I danced with Peter all last night. - Betty: Did you? Aren't you tired of _____ with him? I think he _____ like an elephant.

Mary: Well, he didn't _____ so badly last night.

Uppgift 2: Här är de fyra orden att välja bland för den här uppgiften:

a.	*b.*	*c.*	*d.*
drink	*drinks*	*drinking*	*is drinking*

It's very cold in Scotland. Mack must _____ a cup of hot tea to keep warm. He is fond of _____ hot tea when it's cold, and he _____ many cups every day. In this picture we see him when he _____ tea in the Highlands. He never goes to bed without _____ at least two cups of tea. He likes to _____ his tea with a lot of sugar in it, but his mother says he'll get fat from _____ so much tea with sugar in it. She herself only _____ one cup a day and without sugar.

VÄND BLAD OCH FORTSÄTT DÄR !

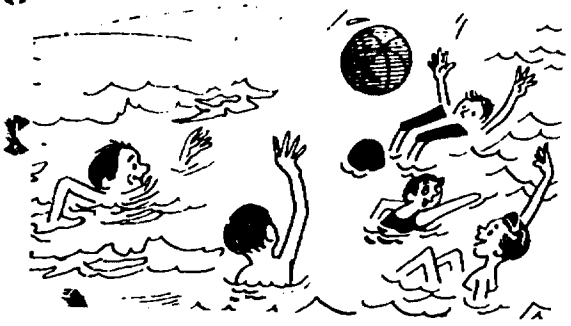
Uppgift 3: Här är de fyra orden att välja bland för den här uppgiften:

a.	b.	c.	d.
watch	*watches*	*watching*	*is watching* *are watching*

These monkeys _____ TV. They are very interested in _____ cowboy films and they _____ at least five such films every week. Daddy Monk says his children learn a lot by _____ TV, but I think he says that because he likes to _____ TV himself. He often _____ TV for hours after all the baby monks are in bed. But don´t tell Mummy Monk for she thinks he gets tired from _____ TV so much. She thinks he is out gathering bananas while really he _____ TV.

Uppgift 4: Här är de fyra orden att välja bland för den här uppgiften:

a.	b.	c.	d.
play	*plays*	*playing*	*is playing* *are playing*



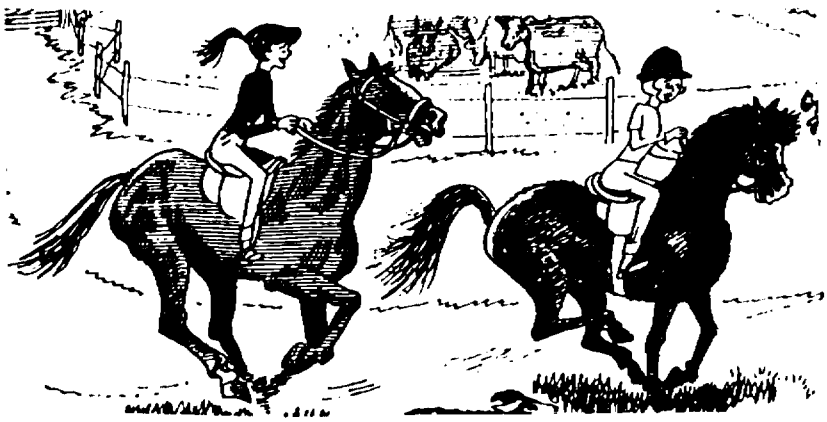
Peter and John _____ water-polo every day in summer. Here they _____ with some friends. John sometimes _____ with his brother, but he is not so good at _____ it as John is.

Here Kate _____ with her cat. She likes to _____ with it very much, and she _____ with it every day. She doesn´t seem to get tired of _____ with it.

VÄND BLAD OCH FORTSÄTT DÄR !

Uppgift 5: Här är de fyra orden att välja bland för den här uppgiften:

a.	b.	c.	d.
ride	*rides*	*riding*	*is riding*



Ann usually _____ her horse in the mornings. In this picture Pat _____ the farmer's brown horse. Sam is very bad at _____, but Pat can _____ like a cowboy. Here she _____ across the fields together with Ann who also _____ quite well.

Uppgift 6: Här är de fyra orden att välja bland för den här uppgiften:

a.	b.	c.	d.
read	*reads*	*reading*	*am reading*

I am very interested in _____. - Oh, are you? - Yes, I'm very fond of _____ about animals, and I _____ a book about tigers just now.

Uppgift 7: Här är de fyra orden att välja bland för den här uppgiften:

a.	b.	c.	d.
put	*puts*	*putting*	*is putting*

He went to bed without _____ out the light. He never _____ out the light. He always forgets to _____ it out. You can save electricity by _____ out the light.

VÄND NU INTE BLAD FÖRRÄN DU BLIR TILLSAGD !

DELPROV D

Här är tjugo meningar. I varje mening har ett ord fallit bort. Det står till höger om sin mening. Du skall nu sätta in det på rätt plats. Markera med ett kraftigt lodrätt streck var du tycker att ordet skall stå. Gör så här:

Mr Smith is a teacher.	not
1. Grandmother has been to Brighton.	never
2. I don't understand why he remembers.	never
3. It's a fact that he comes home before 7.	seldom
4. Do you know why Kate wears mini skirts?	never
5. She comes home late.	never
6. Mr MacFee is late in the mornings.	always
7. Susan practises the piano on Sundays.	never
8. When he is at home, he wears his suit.	seldom
9. It's not true that he gives away money.	seldom
10. He doesn't like TV very much but he watches it in the evenings.	often
11. You know that I try to do my best.	always
12. I came home late when I was at school.	never
13. Mr Austin is fond of smoking and smokes a pipe after breakfast.	often
14. He works hard at school.	never
15. They go for walks on Sundays.	always
16. He goes to dances in winter.	seldom
17. Nowadays he plays with his children.	always
18. In winter he feels tired.	often
19. In spring it is difficult to work hard.	often
20. I don't think it's true that he reads his homework.	never

DELPROV E

Nu tänker vi oss att du är nyfiken. Varje gång jag talar om en sak för dig, så vill du veta mer och du ställer därför en fråga med samma verb i.
Exempel: He is from Gothenburg. *Is he from Hisingen then?*

1. Peter likes tea very much.	*coffee too* ?
2. Mary worked in London last year.	*in a shop* ?
3. He rides like a cowboy.	*every day then* ?
4. He is smoking now.	*a cigar* ?
5. He did it this morning.	*How* *it* ?
6. I get up very early.	*And Sam,* *early, too?*
7. He plays the guitar.	*In which band* ?
8. I watched TV yesterday.	*What programmes* ?
9. He often comes home late.	*When* *home then* ?
10. I was in Finland last summer.	*in Åbo, too* ?
11. You are very sweet.	*as sweet as Mary* ?
12. He speaks many languages.	*German, too?*
13. I drink a lot of tea nowadays.	*it with milk* ?
14. I watched the show there.	*And they,* *it, too* ?
15. I drink milk every morning.	*And Betty,* *, too* ?

VÄND INTE BLAD FÖRRÄN DU BLIR TILLSAGD !

GUME Ld-IC - 3/70.

DELPROV F

I var och en av följande meningar finns en lucka. Du skall sätta in some, any, somebody, anybody, something, eller anything. I stället för att skriva ut orden sätter du ett kryss i rätt ruta till höger.

	some	any	somebody	anybody	something	anything
1. How could ____ believe what he said?						
2. Don't forget to write ____ letters!						
3. Have you ____ cats?						
4. I want an orange. Have you got ____ ?						
5. Would you like ____ apples ?						
6. Why don't you do ____ about it?						
7. Are there ____ pictures in the book?						
8. Did you find ____ money in the box?						
9. I don't know ____ in London.						
10. They can't find ____ shoes in there.						
11. It could happen to ____ .						
12. I never have ____ money on me.						
13. ____ can make a mistake.						
14. Did I tell you that ____ just called?						
15. It's very easy, ____ child could do it.						
16. He left without saying ____ .						
17. I think ____ told me I couldn't do it.						
18. There are ____ people who don't like fish.						
19. I think Tom knows ____ about it.						
20. You may say ____ you like.						
21. There is ____ I don't like in that story.						
22. He went away without saying a word to ____ .						
23. Couldn't you give me ____ icecream?						
24. There is ____ about him I don't like.						
25. They spoke English without ____ accent.						

	some	any	somebody	anybody	something	an
26. I´m not sure, but I think he would like ____ more coffee.						
27. The doctor couldn´t say ____ about the patient.						
28. They feel they must have ____ to read during the week-end.						
29. John: What do you want to do? Mary: ____ you say! It doesn´t matter!$						
30. This car isn´t very expensive, is it? No, it isn´t. ____ can buy it.						
31. He laughed very much at the story. Well, he will laugh at ____ .						
32. He hasn´t had any food for days. What would he like? - ____ will do.						
33. He is faster than ____ else in his class.						
34. I have given him much help. - Why don´t you give him ____ money, too?						
35. It looks as if they never sell ____ in that shop.						
36. You don´t have to ask an expert. You can ask just ____ .						
37. He seldom puts ____ butter on his bread.						
38. When I come home for dinner, there is seldom ____ left for me.						
39. The cake tastes very nice, but I don´t want ____ more, thank you.						
40. I´m glad you liked it. Would you like ____ more soup?						

DELPROV G

Uppgifterna här nedan består av två meningar, en fråga och ett svar. I svaren fattas ett par ord. Det är dem du skall fylla i på de tomma raderna. När du svarar skall du hela tiden göra klart för den som frågar att han har rätt beträffande första delen av sin fråga men fel beträffande den andra delen.

Exempel: I suppose he has long hair and is very short?

Well, he has long hair, but he *is not* *very short.*

1. I suppose you like cocoa and drink it every day?

Well, I like cocoa, but I _____ *it every day.*

2. She is Italian and has lived in Rome, I think?

Well, she is Italian, but she _____ *in Rome.*

3. They are clowns and come from Russia I suppose?

Well, they are clowns, but they _____ *from Russia.*

4. I suppose you heard all the questions and answered them correctly?

Well, I heard the questions, but I _____ *them correctly.*

5. I surpose Mr Austin has a car and washes it every week?

Well, he has a car, but he _____ *it every week.*

6. She sat down and phoned her doctor at once I suppose?

Well, she sat down, but she _____ *her doctor.*

7. She goes to the hospital and plays with the children I believe?

Well, she goes to the hospital, but she _____ *with the children.*

8. I suppose Sam likes his teacher and talks about him very often?

Well, he likes his teacher, but he _____ *about him.*

9. Mr Brown is very rich and buys a new car every year I believe?

Well, he is very rich, but he _____ *a new car every year.*

10. I suppose I am so brown that I look like an Indian?

Well, you are brown, but you _____ *like an Indian.*

11. I suppose he talked a lot but worked at the same time?

Well, he talked a lot, but he _____ *at the same time.*

12. I suppose he comes home very late, and watches TV?

Well, he comes home late, but he _____ *TV.*

13. I suppose a pelican is a bird that eats other birds?

Well, it is a bird, but it _____ *other birds.*

14. I suppose you came home early because you were tired?

Well, I came home early, but I _____ *tired.*

15. He took it back, and then he did it himself, I believe?

Well, he took it back, but he _____ *it himself.*

VÄND INTE BLAD FÖRRÄN DU BLIR TILLSAGD !

Appendix B

THE PUPIL ATTITUDE TEST

Lärarhögskolan i Göteborg
GUME-projektet
Ld-IC 5/70

Namn: _____ Klass: _____

Skola: _____

Engelsklärare: _____

A) Intresse för olika skolämnen.

Du skall här få tala om vad du tycker om de olika ämnen som ni har i skolan i år. Du skall göra det genom att sätta ett kryss (x) för varje ämne inom parentesen under den pil som bäst visar hur du tycker om ämnet. Du skall inte tala om vad du tyckt nu under de senaste veckorna när ni haft Gume- engelska utan hur det var före och hur det är i vanliga fall, hur du brukar tycka när allt är som vanligt. Hoppa inte över något ämne!

	Nästan alltid roligt ↓	Mera roligt än tråkigt ↓	Mera tråkigt än roligt ↓	Nästan alltid tråkigt ↓
Svenska	( )	( )	( )	( )
Matematik	( )	( )	( )	( )
Engelska	( )	( )	( )	( )
Kristendomskunskap	( )	( )	( )	( )
Samhällskunskap	( )	( )	( )	( )
Historia	( )	( )	( )	( )
Geografi	( )	( )	( )	( )
Naturkunskap	( )	( )	( )	( )
Musik	( )	( )	( )	( )
Teckning	( )	( )	( )	( )
Slöjd	( )	( )	( )	( )
Gymnastik	( )	( )	( )	( )

B. Elevenkät

Vi vill nu veta lite grand om vad du tyckte om Gume-projektet.

Svara med (x) eller korta meningar.

1. Det som var bra med GUME-lektionerna var att

_____

_____

_____

2. Det som inte var bra med GUME-lektionerna var att

_____

_____

_____

3. På de här timmarna tyckte jag att jag lärde mig

_____ väldigt mycket

_____ rätt så mycket

_____ så där lagom

_____ ganska litet

_____ väldigt litet

4. De här timmarna var

_____ väldigt roliga

_____ ganska roliga

_____ så där lagom

_____ ganska tråkiga

_____ väldigt tråkiga

5. När vi gjorde muntliga och skriftliga övningar så förstod jag vad det var vi höll på med och vad man skulle göra

_____ alltid

_____ för det mesta

_____ibland

_____ganska sällan

I GUME-projektet har alla klasserna fått lära sig samma saker fast på olika sätt, vi har använt olika metoder. Försök nu tala om vad du tycker om den metod du hade i din klass (när vi här talar om förklaringar så menar vi inte förklaringar om att du skulle vända blad, var du skulle titta osv. utan grammatiska förklaringar, där vi försökte tala om vad vi vid varje tillfälle höll på att öva och varför man säger så på engelska.)

6. a) _____ i min klass fick vi grammatiska förklaringar

b) _____ i min klass fick vi inte grammatiska förklaringar

Om du kryssat för a) här ovan så gå vidare till fråga 7, om du kryssat för b) så gå i stället över till fråga 8 direkt.

7. (denna fråga skall endast besvaras av dem som valt 6 a ovan)

A. i min klass fick vi förklaringar på svenska, men det hade varit bättre om vi hade fått dem på engelska: ja eller nej? _____

i min klass fick vi förklaringar på engelska, men det hade varit bättre om vi hade fått dem på svenska: ja eller nej? _____

B. Jag tycker att förklaringarna

_____ gjorde det mycket lättare att förstå

_____ gjorde det något lättare att förstå

_____ inte gjorde någon skillnad

_____ gjorde det något svårare att förstå

_____ gjorde det mycket svårare att förstå

C. Jag tycker att vi fick

_____ alldeles för lite förklaringar

_____ något för lite förklaringar

_____ lagom mycket förklaringar

_____ något för mycket förklaringar

_____ alldeles för mycket förklaringar

8. (denna fråga skall bara besvaras av dem som svarade med alternativ b) i fråga 6 och som alltså hoppat över fråga 7)

_____ Jag saknade inte förklaringarna och tyckte inte jag behövde några sådana.

_____ Jag saknade förklaringar ibland och tror att det varit bra med en del sådana.

_____ Jag saknade förklaringar rätt ofta och skulle velat ha sådana rätt många gånger.

_____ Jag saknade förklaringar väldig mycket och skulle velat ha det ofta.

I GUME-lektionerna förekom regelbundet s.k. fyrfasövningar(vi ställde en fråga på bandet, ni fick besvara den, så kom rätt svar på bandet och ni upprepade det i kör). Du skall nu i fyra frågor tala om vad du tyckte om dessa övningar.

9. Jag tyckte att jag i fyrfasövningarna lärde mig att <u>tala engelska</u>

_____ väldigt bra

_____ ganska bra

_____ så där lagom

_____ ganska litet

_____ väldigt litet

10. Jag tyckte att jag i fyrfasövningarna lärde mig <u>engelsk grammatik</u> (hur man skall säga för att det skall bli riktig engelska).

_____ väldigt bra

_____ ganska bra

_____ så där lagom

_____ ganska litet

_____ väldigt litet

11. Jag tyckte att fyrfasövningarna var

_____ väldigt roliga

_____ ganska roliga

_____så där lagom

_____ganska tråkiga

_____väldigt tråkiga

12. Jag tyckte att fyrfasövningarna var

_____ väldigt lätta

_____ ganska lätta

_____så där lagom

_____ ganska svåra

_____ väldigt svåra

Appendix C

PUPILS' INTEREST IN VARIOUS SCHOOL SUBJECTS

Pupils' Interest in Their School Subjects.

Table C-1 gives the means calculated from the 4-point scale used to investigate the pupils' interest in the 12 subjects they take in the 6th form. Means have also been calculated per subject (vertically) and per class (horizontally).

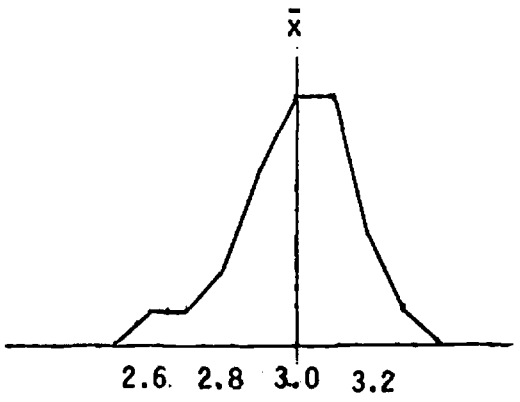
The various subjects rank as follows:

3.5	Drawing, Handicraft, Gymnastics
3.4	
3.3	
3.2	Geography
3.1	Science
3.0	Mathematics, English, History
2.9	
2.8	Civics
2.7	
2.6	Swedish, Music
2.5	
2.4	
2.3	Religion

This survey spekas for itself. It should be noticed that 3.0 corresponds to "more fun than boring" and 2.0 to "more boring than fun".

The class means vary in the following way:

	Im	Ee	Es	All
3.3	1			1
3.2	1	1	1	3
3.1	4	1	2	7
3.0	1	3	3	7
2.9	1	3	1	5
2.8	1	1		2
2.7			1	1
2.6			1	1
means	3.1	3.0	2.9	3.0



The distribution of the figures is normal and the difference between the methods small.

Table: C-1.

Class	Swedish	Mathematics	English	Religion	Civics	History	Geography	Science	Music	Drawing	Handicraft	Gymnastics	Means per class
1	2.6	3.0	3.2	2.4	3.2	3.3	3.2	3.4	2.1	3.5	3.6	3.5	3.1
2	2.1	3.2	3.0	2.2	2.5	2.8	3.2	3.4	2.0	3.2	3.5	3.8	2.9
3	2.8	3.5	3.0	1.8	3.6	3.1	3.2	3.0	2.4	3.2	3.6	3.3	3.0
4	2.4	4.1	2.8	2.5	2.7	3.3	3.7	3.2	1.8	3.6	3.8	3.8	3.1
5	3.3	2.8	3.7	2.7	3.0	3.1	3.5	3.3	2.5	3.6	3.6	3.4	3.2
6	2.1	2.6	3.1	2.3	4.1	3.2	3.1	2.9	2.8	3.7	3.2	3.6	3.1
7	2.0	2.8	2.9	2.4	2.4	2.1	3.5	3.1	2.5	3.3	3.1	3.0	2.8
8	2.7	3.0	3.4	2.9	2.9	3.1	3.2	3.4	3.5	3.6	3.8	3.5	3.3
9	2.2	3.4	2.9	2.5	2.6	3.7	3.6	3.0	2.2	3.8	3.9	3.7	3.1
10	2.9	3.0	2.8	2.6	2.5	3.3	3.1	2.9	2.4	3.6	3.7	3.7	3.0
11	2.8	2.8	3.2	2.5	2.4	3.2	3.6	3.0	2.9	2.5	3.5	3.2	3.0
12	2.6	2.8	3.1	2.0	2.7	2.5	3.7	3.4	2.0	3.4	3.3	3.2	2.9
13	3.1	3.2	3.4	2.3	2.9	3.5	3.4	3.4	3.2	3.7	3.1	3.7	3.2
14	2.4	3.0	2.5	1.5	2.0	2.6	3.6	3.1	2.3	3.8	3.7	3.4	2.8
15	2.2	3.1	2.9	1.9	2.7	2.8	3.1	2.6	3.3	3.4	3.3	3.4	2.9
16	2.8	2.7	3.3	2.5	2.8	3.2	3.1	3.2	3.3	3.8	3.1	3.7	3.1
17	2.8	3.0	3.2	1.6	1.8	3.0	3.5	3.1	3.0	3.4	3.3	3.4	2.9
18	2.4	3.4	2.9	1.9	2.8	3.1	2.7	3.0	2.7	3.5	3.8	3.6	3.0
19	2.0	2.8	2.4	2.1	2.8	2.1	2.6	3.1	1.9	3.5	2.9	3.8	2.7
20	2.6	2.7	2.5	1.7	2.8	3.6	3.8	3.3	2.2	3.5	2.9	3.3	2.9
21	2.6	2.7	2.7	2.9	2.6	3.2	3.3	2.9	2.7	2.8	3.3	3.9	3.0
22	2.7	2.7	3.4	2.2	3.4	2.8	2.7	3.4	2.9	3.7	3.4	3.3	3.1
23	2.5	3.2	3.3	2.8	2.6	3.1	3.4	3.0	3.6	3.6	3.4	3.4	3.2
24	2.5	2.6	2.9	2.4	2.6	2.6	3.3	3.0	2.8	3.7	3.8	3.6	3.0
25	2.8	3.2	3.0	2.8	2.8	3.0	2.7	2.9	2.6	3.5	3.6	3.1	3.0
26	2.2	2.7	2.4	1.6	2.2	2.9	2.6	2.6	2.1	3.3	3.9	2.9	2.6
27	2.7	3.2	2.9	2.8	2.8	3.0	3.2	3.3	3.1	3.6	3.7	3.3	3.1
Means per Subject	2.6	3.0	3.0	2.3	2.8	3.0	3.2	3.1	2.6	3.5	3.5	3.5	

It should be stressed that these figures are means of means, and a difference of .7 as between classes 8 and 26 is thus a very great one, indicating that the teaching climate in the two classes is completely different.

The overall means for all subjects per method in the project are: Im 3.1, Ee 3.0, Es 2.9.

The lowest figures, below 2.0, almost all occur in the Religion column, a subject which obviously does not appeal to the pupils. No class has a mean higher than 2.9 in Religion. The subject receiving the most varied ratings seems to be Music. There are two low figures of 1.8 and 1.9, but there are also high figures of 3.6 and 3.5. It seems likely that the teacher factor makes itself felt strongly in a subject like Music where the teachers no doubt represent very different degrees of proficiency themselves.

In Civics the figures vary between 1.8 and 4.1, an even wider gap for similar reasons probably. Some teachers make the subject more interesting than others.

It is beyond the scope of this project to relate pupils' interest to a possible teacher variable. Nor have correlations been calculated for the relationship between interest and grades, or interest and social class. An inspection of the figures compared to geographical location of the schools does not yield any clear-cut results. Some classes that might have been expected to show negative attitudes to school do so, others do not, and vice versa.

Appendix D

THE TEACHER ATTITUDE TEST

Lärarhögskolan i Göteborg
GUME-projektet
Ld - IC - 5/70

LÄRARENKÄT - GUME IV

1. Namn: _____

2. Min klass hade _____-metoden under försöket (Im, Ee eller Es).

3. Jag brukar nog själv i åk 6 följa vad som närmast torde motsvara
   ___ Im (i princip inga teoretiska färklaringar, hela undervisningen på eng.)
   ___ Ee (hela undervisn. på eng., grammatiska kommentarer till vad som övas)
   ___ Es (gramm. kommentarer och förklaringar på svenska och jämförelser med svenska där så är lämpligt)

4. Jag brukar ge grammatiska förklaringar (på svenska eller engelska)
   ___ varje lektion
   ___ rätt ofta och regelbundet
   ___ någon gång ibland när det är nödvändigt
   ___ sällan eller aldrig

5. När vi har förklaringar brukar jag
   ___ själv ge dem snabbt och koncist
   ___ låta någon elev ge dem och ev. själv runda av efteråt

6. Mina lektioner brukar nog i allmänhet vara till ca _____ % på engelska.

7. Jag brukar använda muntliga strukturövningar liknande dem som förekom i GUME-lektionerna
   ___ alltid när vi övar nya grammatiska moment
   ___ rätt ofta och regelbundet
   ___ någon gång ibland
   ___ aldrig

Som stöd för minnet vid besvarandet av de följande frågorna: Vi övade följande moment under projektet: s-formen, do-konstruktionen, adverbplaceringen, some-any, prep+ing-form, presens-progressiv form, imperfektum.

8. Bra med den metodik som min klass undervisades efter var:

_____

_____

_____

_____

Mindre bra eller dåligt var:

_____

(9. - forts) _____

_____

10. Ange kortfattat Din åsikt om

a) De muntliga övningarna:

_____

_____

_____

b) De skriftliga övningarna:

_____

_____

_____

c) Lästexterna:

_____

_____

_____

d) (för E-grupperna:) Förklaringarna (dela gärna upp på de olika gramm. momenten som ingick, se ovan)

(för Im-grupperna:) Avsaknaden av förklaringar (när saknades de mest, hur tror Du att eleverna upplevde detta etc)

_____

_____

_____

_____

11. Om tempot i lektionerna - pauslängder och talhastighet - anser jag:

_____

_____

12. Om den tekniska kvaliteten på materialet (band, texter, overheadblad) anser jag

_____

_____

13. I projektet användes bandspelare, högtalare, overheadprojektor. Vad anser Du om denna tekniska materiel (fungerade den bra, innebar den extra arbete etc?):

14. Jag har gjort följande iakttagelser jämfört med vanlig undervisning beträffande

a) elevernas intresse: _____

_____

b) disciplinen i klassen: _____

_____

c) inlärningseffekten (subjektivt bedömd innan provresultaten föreligger):

_____

_____

15. Om provet (som gavs som för- och efterprov) anser jag:

_____

_____

16. Kommentarer - positiva och negativa - till de enskilda lektionerna (gärna lektionsvis för alla tolv; fortsätt på baksidan om det behövs!):

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

17. På det hela taget tycker jag att den tid som experimentet tagit varit

___ i det närmaste helt bortkastad
___ tämligen outnyttjad
xxxxxxxxxxxxxxxxxxxxxxxx
___ tämligen väl utnyttjad
___ mycket väl utnyttjad

Ytterligare kommentarer:

Appendix E

PARTICIPATING TEACHERS

Participating Teachers (in alphabetical order).

Name:	School:
Monika Ahlberg	Ekebäck
Lilian Ahlbäck	Järnbrott
Inga-Lill Alvarsson	Högsbo
Gunn Augustsson	Kyrkbyn
Lars Bergsten	Jättesten
Georg Blom	Gamla Lunden
Vivi-Anne Blomberg	Flatås
Marita Carlsson	Järnbrott
Birgit Ferm-Karlsson	Kannebäck
Barbro Forkby	Ekebäck
Åke Hallén	Kyrkbyn
Ingegärd Holger	Ekebäck
Monica Karlberg	Högsbo
Gunnar Linde	Bjurslätt
Ulrika Linderum	Svartedalen
Olle Nyqvist	Dala
Ann-Christin Persson	Bjurslätt
Ebba Petersson	Jättesten
Ulla du Rietz	Kyrkbyn
Ann-Sofie Runmalm	Flatås
Elisabeth Rylander	Flatås
Birgitta Sandén	Guldheden
Bo Sibbesson	Bjurslätt
Anita Sidén	Tynnered
Margot Starzmann	Kannebäck
Birgitta Stengård	Guldheden
Sven Wirén	Järnbrott

Appendix F

DESCRIPTIVE STATISTICS PER SCHOOL CLASS (N = 27)

AND FOR THE TOTAL EXPERIMENTAL POPULATION (N = 577)

Means per School Class (N = 27)

School class no.	N	IQ	Grades	Std Test	PACT	Pre-test	Post-test	Pro-gress	Pupil Attitudes
01	19	51.00	27.63	46.89	33.00	45.95	59.74	13.79	24.53
02	26	56.73	28.50	56.35	36.00	54.88	74.08	19.19	22.17
03	19	56.71	28.58	61.58	34.42	66.42	84.00	20.72	20.63
04	23	53.00	26.55	45.61	31.74	38.61	52.87	14.26	22.71
05	18	50.78	26.50	46.61	34.29	42.28	61.89	19.61	26.22
06	19	56.68	30.95	61.53	37.79	61.05	81.58	20.53	24.11
07	22	52.79	25.77	47.00	33.06	40.95	53.50	12.55	22.77
08	22	50.27	26.86	42.68	26.05	47.68	57.64	9.95	21.17
09	13	49.25	30.23	49.23	37.69	45.46	66.54	21.08	23.67
IM	181	53.26	27.84	50.94	33.55	49.24	65.35	16.52	23.01
10	18	52.89	27.67	45.44	30.00	41.94	53.83	11.89	21.27
11	22	52.48	28.23	51.55	33.82	48.32	66.59	18.27	24.00
12	24	55.42	27.50	66.54	36.71	64.17	81.00	17.22	24.13
13	22	55.68	31.64	62.55	36.64	62.23	82.73	20.50	26.35
14	22	55.77	29.18	64.23	36.59	60.82	76.00	15.18	19.11
15	22	59.27	30.82	56.86	35.57	51.32	70.09	18.77	24.60
16	20	51.45	27.45	49.35	32.33	48.75	64.20	15.45	26.00
17	25	53.08	26.25	61.36	37.44	56.96	82.84	25.88	23.86
18	20	51.40	24.75	40.85	29.74	39.00	51.85	12.85	22.11
Ee	195	54.25	28.19	56.04	34.61	53.14	70.79	17.64	23.53

School class no.	N	IQ	Grades	Std Test	PACT	Pre-test	Post-test	Pro-gress	Pupil Attitudes
19	24	52.21	30.25	45.13	30.29	43.75	56.54	12.79	18.64
20	22	53.23	27.00	53.10	38.85	56.00	76.09	20.09	20.81
21	19	57.00	31.26	64.89	39.67	61.32	85.53	24.21	22.88
22	22	56.23	28.91	61.09	35.67	62.50	82.50	20.00	23.10
23	22	59.73	28.77	62.14	37.14	61.68	78.68	17.00	26.50
24	23	51.18	27.91	50.39	34.48	49.70	64.96	15.26	20.65
25	26	51.42	27.00	56.85	35.13	49.73	68.12	18.38	23.52
25	16	52.50	22.13	43.75	28.87	41.75	54.81	13.06	20.15
27	27	48.81	26.33	43.33	30.85	45.08	60.81	17.23	23.76
Es	201	53.43	27.82	53.20	34.63	52.27	69.58	17.54	22.34
♂	275	53.91	26.69	49.95	33.60	49.26	64.07	15.21	22.04
♀	302	53.43	29.11	56.73	34.91	53.74	72.84	19.11	23.70
Ak 6	577	53.66	27.95	53.48	34.29	51.61	68.67	17.26	22.94

Means and Standard Deviations for the Total Population (N = 5)

	$\bar{x}$	s	N
IQ Verbal	5.30	1.79	564
IQ Inductive	5.79	1.93	564
IQ Spatial	5.56	1.97	564
IQ Total	53.66	9.64	564
Grades Swedish	3.15	.92	573
Grades English	3.09	1.03	576
Grades Maths	3.08	.97	576
Grades Total	27.95	7.71	573
Std Test EL	12.09	4.81	569
Std Test EM	12.48	6.31	569
Std Test EA	15.94	5.47	569
Std Test EU	12.96	5.22	569
Std Test Total	53.48	18.68	569
PACT	34.29	8.77	550
Pre-test	51.61	20.89	575
Post-test	68.67	27.16	576
Progress	17.26	12.32	574
Pupil Attitudes	22.94	4.41	529