ABSTRACT
                This collection of papers assesses the validity of
the MLAT. It is not self-evident that a test designed for English
speaking university students in America and validated there is
necessarily valid in an English university. The differences between
American English and British English are evident, and American
universities admit a greater proportion of the population than do
English ones and different entry qualifications apply. It is also
possible that different educational methods lead to the strengthening
of different aptitudes. The collection consists of three main
sections: (1) Descriptions of the experiment and discussion of the
MLAT with particular reference to its internal and predictive
validity. A multiple regression analysis is used for this. (Culhane).
(2) A description of the performance of a select group of students of
"proved" linguistic ability, including error analysis of the
sub-tests and a description of learning strategies for Part V. (Lee)
(3) Detailed item analysis of sub-tests III and IV to determine which
were the relevant factors operating in a speeded and a non-speeded
sub-test. There is also a limited assessment of the validity of Part
V in relation to a memory-based course in Scientific Russian.
(Shepherd and Alford). (CHK)

University of Essex

Language Centre

CASIONAL PAPERS

ter

ster, November 1970

7

L2   LEARNING

[1]Lecturer in the Language Centre

[2]Lecturer at the Institut de Phonétique, Université
      Libre de Bruxelles

[3]Lecturer in English at Tunghai University, Taiwan

[4]Formerly Fellow of the Language Centre

Introduction

        The idea of assessing the validity of the
Carroll - Sapon Modern Language Aptitude Test was
conceived by Professor Peter Strevens soon after
his appointment as Director of the Language Centre
of this University. It is not self-evident that a
test designed for English-speaking university
students in America and validated there is necess-
arily valid in an English university situation.
The differences between American English and our
own variety are obvious, especially differences in
pronunciation. Secondly the American university
tends to take in a greater proportion of the popu-
lation than in this country. Thirdly the education
of the American pupil may be vastly different from
that of his English counterpart. Different methods
may have been used leading to a strengthening of
different aptitudes.

        The Language Centre has a wide variety of
courses available. Here is a list of the main ones:
    1)   Preliminary Language Year (PLY)

        This is a course in the School of Compara-
tive Studies for students who intend to specialise
in Government, Literature, Art, Sociology or
Language (starting 1971). It is a one year intensive
course in Portuguese, Russian, or Spanish. Students
who learn Russian will later specialise in the
Soviet Area, and those who study Spanish or Portu-
guese will specialise in the Latin American Area.

    2)   Common First Year (CFY) Russian

        This is a continuation of PLY. By the
end of this year students will be expected to be
able to read Russian texts in the original. Students

who have a good A-level in Russian are not expected
to undertake an intensive course, but are admitted
directly into CFY.

3)   CFY Spanish and Portuguese "Conversion"
      Courses

Students who had earlier specialised in
Spanish "convert" to Portuguese, and those who
specialised in Portuguese "convert" to Spanish. As
in the case of Russian, students with a good A-level
pass are admitted directly to CFY.

4)   CFY French

Students who choose to follow a North
American Option are expected to study French in
their first year. This is the largest single
group of students, all of whom are required to
have at least O-level French.

5)   Social Studies French, Russian,
      Portuguese, Spanish.

These are "service" courses for other
departments, normally of two years duration. No
previous knowledge of the language is necessary.

6)   M.A. in Applied Linguistics

This is a course for "in-service" teachers
who wish to improve their qualifications.

There have been other courses taught by staff
of the Language Centre, some of which no longer
function (for example, Russian and German for
Scientists) and some of which have been started
much more recently, e.g. M.A. in Linguistics, and
Elementary Italian. None of the students from
these two categories are covered in the main part
of the investigation. The section by Mark Alford
is concerned with scientists learning Russian, but
this only applies to Part V of the Test (Paired
Associates).

The Test was performed in a language labora-
tory, so that conditions were as near as possible
identical for all the students taking the Test.
Timing was done by master tape, which had to be
turned over at the halfway stage. The invigilator
also had to stop the tape at the end of Part I
(Number Learning) for the students to transfer their
answers to a form suitable for automatic marking.
Thanks are due to the members of staff of the

5

Language Centre who helped with invigilation and marking.

In order to carry out the multiple regression, the program XDS3 was used. Mr. Ian Russell of the ·Department of Mathematics adapted the program to our needs and helped enormously in interpreting the data. He also contrived to get more information from the program than had so far been achieved in this university, in that he used it to obtain factor loadings. I have included in the text those statistics which were more readily understandable to the layman. For those who are more "statistically minded" fuller details are included in the Appendix at the end of the Occasional Paper.

It is a difficult task to measure aptitude and to isolate factors for an aptitude test, hence come of the terms used in Carroll's Factor Analysis are somewhat vague, e.g. "sound-symbol association ability", "associative memory". It is also difficult to know how aptitude develops when it is exercised by the right sort of study. If it does develop, then surely it becomes more difficult to discern innate aptitude and to discriminate between the aptitudes of pupils of different environmental backgrounds.

Notwithstanding all the difficulties involved in the overall problem, we hope that this work adds to the body of knowledge and that people concerned with language testing will find it useful.

PTC

iv

Acknowledgements

Some Statistical Terms in Simple Language for Linguists

P. T. Culhane

Mean Score	Average score.
Median Score	Score at the centre point on the rank order of results where half of those being tested fall on each side of the line.
Deviate	The amount by which each score exceeds or falls short of the mean.
Variance	The sum of the squares of the deviates.
Standard Deviation	Obtained by taking the square root of the Variance. Used to show how well the test 'spreads' the population being tested.
Histogram	A pictorial method of demonstrating distribution. Used to show how many students obtained how many marks in a particular test or subtest.
Mode	The highest point in a histogram. The most frequent value in a distribution curve or histogram.
Percentile	A method of demonstrating the level of performance of an individual in relation to a control or 'norm' group. Thus, if a student is on the 90th percentile, 10% of the students in the norm group would have been above him, and 90% lower. An advantage of this method is that individuals are not penalised for

being low in the rank order of their own group.

Correlation

A method of comparing one test with another in numerical terms. A correlation of 0 means that the tests are measuring different things. A correlation of 1 would mean that they are measuring the same things. A minus correlation would cast suspicion on the internal validity of the tests being compared.

Factor

One of the components being measured by a test or subtest. Factors may be isolated by statistical means and given a weighting relative to other factors in a particular test.

Variable

Anything which discriminates between one student and another e.g. sex, school, year, course of study, previous language study, exam mark etc. etc. In this investigation there are nineteen variables, which fall into different categories (see below).

Multiple Regression

A method of producing a prediction equation by comparing variables. The primary aim of MLAT is to predict. The object of this investigation is to assess how well MLAT has operated in the University of Essex. We shall do this by seeing how much the individual parts of MLAT contribute to the prediction of course mark and examination mark and how a prediction with MLAT compares with predictions we can make using previous language study on its own.

Description of MLAT and its uses at Essex

P. T. Culhane

What is MLAT?

During this description I shall refer to the
Manual for the Test (No. 5 in the bibliography) and
to the Factor Analysis produced by John B. Carroll
in the Journal of General Psychology, 59, pp 3-19
(No. 1 in the bibliography).

"The MLAT has been designed chiefly to
provide an indication of an individual's
probable degree of success in learning a
foreign language, but it is also useful in
predicting success in learning to read write
and translate a foreign language. It is
applicable in connection with both 'modern'
spoken languages and ancient languages such
as Latin or Greek" [5].

Each student is given an answer sheet, a Test
Booklet and a Practice Exercise Sheet. The first
two parts of the Test are on tape, and timing and
instructions for all five parts are also done by
the tape.

The test consists of five parts as follows.

Part I   Number Learning

Candidates are taught the numbers 1,2,3,4,10,
20,30,40,100,200,300,400 in a language unfamiliar
to them. They are given an opportunity to practise
the numbers and correct their mistakes. This
provides a kind of short-term reinforcement. The
numbers are then read out at random and the student
has to write them down. When he reaches the end of

10

Part 1 the tape is designed to enable the student to transfer the numbers to a format which may be marked by machine or by means of a specially designed marking sheet with holes.

"This seems to measure one aspect of the memory component of foreign language aptitude and may also have an 'auditory alertness' factor" [5].

Part II   Phonetic Script

A series of sounds is read, which corresponds to groups of phonetic symbols printed on the answer sheet. Each of four choices in five multiple-choice questions is read. At the second reading only one choice is read, and the student has to decide which one. The sounds in each group of five questions are phonetically associated.

"This appears to measure what we have called sound-symbol association ability, that is the ability to learn correspondences between speech sounds and orthographic symbols. It may also measure a sort of memory for speech sounds, and it tends to correlate highly with the ability to mimic speech sounds and sound combination in foreign languages"[5].

Part III   Spelling Clues

This is a multiple-choice test in which the student is asked to decide which of five meanings corresponds to a word which has a disguised spelling.

e.g. luv
A. carry
B. exist
C. affection
D. wash
E. spy

This test is highly speeded and, according to the Manual, tests English vocabulary and the same kind of sound-symbol association as is measured in Part II.

11

Part IV   Words in Sentences

Students are given two sentences. In the first
sentence a word is underlined and in the second
sentence a number of alternatives are underlined.
The student has to choose the alternative in the
second sentence which has the same function as the
underlined word in the first sentence.

    e.g. London is the capital of England.
         He liked to go fishing in Maine.
         A   B   C D   E

"This part is thought to measure sensitivity to
grammatical structure" [5]. As in the case of Part
III, the test is speeded,though not quite to such
an extent.

Part V   Paired Associates

Students learn vocabulary and are then expected
to choose the correct meaning of a foreign word
from five alternatives.

    e.g. mep
                   A. in
                   B. on
                   C. that
                   D. enter
                   E. art

The stimulus for each item is a printed one.

    "This part measures the rote memory
aspect of the learning of foreign languages"[5].

    What is described above is the long form of the
test. The short form consists of Parts III, IV and
V i.e. the parts which do not have aural stimulus.

Factors in learning a foreign language.

    Carroll lists the following seven factors in
foreign language learning, in order of importance.

         1.   Associative memory.
         2.   Linguistic interest.
         3.   Inductive language learning ability.
         4.   Grammatical sensitivity.
         5.   Sound-symbol association.

6.   Verbal knowledge.
7.   Speed of association.   [1]

Carroll's isolation of factors and his weightings are based on a large number of test variables and not solely the parts of MLAT. The tests which became part of the MLAT were among those used by Carroll and the table below is an attempt to represent Carroll's factor loadings in a non-numerical way, illustrating only the parts of MLAT.

Parts of MLAT \ Factors	1	2	3	4	5	6	7
MLAT I	@	-	-	-	-	-	-
MLAT II	0	+	?	-	@	0	-
MLAT III	0	-	-	+	0	@	-
MLAT IV	+	+	-	@	-	-	-
MLAT V	@	-	-	0	-	-	-

@ = most important factor in subtest.

+ = significant factor.

? = factor which plays some part in test.

- = factor with little or no significance.

It may be seen from the above that the fourth factor i.e. 'grammatical sensitivity' may have more importance in MLAT than it had in Carroll's overall analysis. Factor 3 appears to have less importance. The weightings may, of course, be intentional.

Uses of the test at Essex

The primary reason for the use of MLAT at
Essex is to find out whether or not it is valid in
an English university situation. Students who
come to us have usually studied a language at least
to 'O' level, but not always beyond. We wanted to
know how much MLAT could add to a prediction of
their success, and whether or not the test could
reasonably be used as a selection criterion. The
primary reason for the use of the test was therefore
to investigate the predictive validity of the test
itself.

Other uses of the test may be described as
'fringe benefits'. At the beginning of the year
we were able to circulate an early warning of
students who were expected to be weak, so that a
close watch could be kept on their progress during
the first few weeks of an intensive course. MLAT
has also been used, along with other factors, to
determine which 'stream' a student should be in in
French courses. It has also been possible to
advise students how they stand, according to MLAT,
in relation to other students. This is particularly
useful in dealing with mature students who have
little or no formal training in languages and who
feel that they might like to specialise in language
study.

We have not used MLAT as a selection criterion;
it is the object of this investigation to decide
how far we may do this.

14

The Internal Validity of ILAT

P. T. Culhane

The operation performed by a typical test or examination is
threefold: 1)   It provides a rank order i.e. it measures the
level of competence of each student which places him in relation
to other students taking the test.
              2)   It may be used for remedial instruction, by
indicating individual points of weakness.
              3)   It suggests a pass mark, or minimum level of
acceptability.

It has been suggested that 1) above is not 'relevant' to
students and that what we should expect is that students should
attain a self-satisfying level of competence at which point an
examination becomes 'irrelevant' anyway. I do not wish to enter
this argument here. I do wish to point out that the concept of
'rank order' is important in a test which is intended to determine
how well a student may be expected to learn a foreign language.

For the purposes of this investigation we have data from the
following groups of students.

15

Table 1

	Year	F (CFY)	P (PLY)	R (PLY)	R (CFY)	Sp (PLY)	Sp (CFY)	
Comparative Studies	65	33	-	20	3	11	9	76
	66	60	2	12	1	8	10	91
	67	68	8	13	1	6	-	96
Social Studies	65	37	-	-	18	-	21	76
	66	40	-	-	15	-	31	86
	67	36	-	-	9	-	31	76
		274	10	45	47	23	102	501

92 (R)   125 (Sp)

Note F = French   CFY = Common First Year
      P = Portuguese   PLY = Preliminary Language Year
      R = Russian
      Sp = Spanish

Table 2 Numbers of Students in Schools.

Year	Comparative Studies	Social Studies	
65	76	76	152
66	91	86	177
67	96	76	172
	263	238	501

In addition to the data from our own students we have also a limited amount of data from American experiments on ALA (see [1],[2]&[5]). We also have, through the cooperation of Mr. John Smith of the Technological University of Bath, histograms, mean scores and standard deviations on about five hundred observations. I shall make reference to these data in due course.

16

Table 3   Means and Standard Deviations for the test overall and the subtests.

		Mean	Min	Max	Standard Deviation	Median
Full test	(out of 190)	124.42	43	179	22.62	127.83
Short form (113)		83.72	19	111	15.11	-
SUBTEST I	(43)	34.88	0	43	8.83	38.17
II	(30)	25.81	3	30	3.40	28.48
III	(50)	20.58	3	50	8.23	20.38
IV	(45)	25.56	5	41	6.61	25.08
V	(24)	17.57	0	24	5.88	19.03

It is apparent from the Mean Scores that Part III and IV are the most difficult and Part I is the easiest subtest. The large difference between the Mean and the Median Scores in tests I and V is probably a reflection of the fact that maximum possible marks were obtained by a large number of students in these subtests.

Table 4. Percentiles. These are here presented beside those of the groups of 'College Freshmen' in [3].

Percentile	Essex raw scores	American 'College Freshmen' raw scores	
		Men	Women
95	157	154	154
90	151	143	149
80	143	141	138
70	137	133	134
60	132	126	125
50	127	118	119
40	122	111	110
30	115	103	105
20	107	90	94
10	93	83	78
5	81	73	71
No.	501	465	277
Mean	124.4	115.0	113.8
SD	19	26.4	20.6

As we might expect the mean score of our group is somewhat higher than that of the American sample. The reason for this is probably that the American university system tends to recruit a wider section of the population that ours does and that people who enter the Schools of Comparative and Social Studies tend to be people with experience of language and/or people with high aptitude. The fact that the Standard Deviation for our group is lower indicates a bunching at the higher end of the scale.

18

The Histograms

These should be studied in relation to the idea of a
'normal distribution curve'. In a normal distribution one would
expect a small number of people with a low mark, a large number
in the middle ranges and a small number at the top of the scale.
Thus the curve would look something like this:

No. of
Students

Score
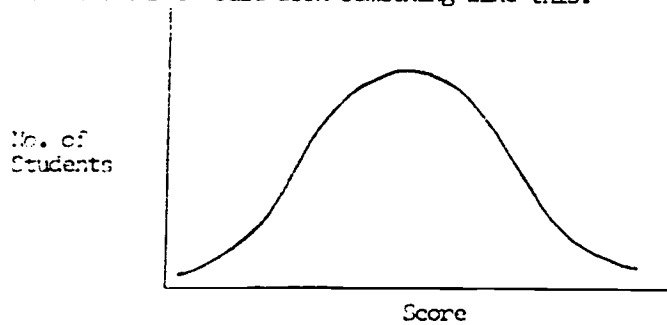
In the histograms the distribution is represented in blocks
of students obtaining marks within a small range. In the case
of Histogram A (see over) the lowest mark obtained by any student
was 43 and this student was the only one to appear in the range
43 - 52. Six people had a mark of between 53 & 62 etc. The Mode
of the histogram is centred on 137·5.

Histograms B C D E & F (see over)

Only the histograms for tests III and IV show anything like
a normal distribution curve. The Mean Scores in these tests are
lower in comparison with the other tests. In the case of Test II
the MODE is not quite on the maximum mark. This might lead to the
speculation that if the test had been tried on a less able
population the Mode would move towards the centre. The mean score
at the University of Bath is 110·3 i.e. 14·1 lower than that at
Essex. There are reasons why one might expect this to be the
case. Bath specialises in Science and Technology and its students
are less likely to have experience of Languages. Secondly, there
is a tendency for students to be accepted into Science Faculties
with lower 'A' level qualifications than those expected of 'Arts'
students. There may be an I.Q. difference. These reasons are
speculative and are not supported by definite evidence. Roy Cox
[15] observes that students in Comparative Studies at Essex are
less well able to perform in reasoning tests than are Social
Scientists and Physical Scientists. However this may be, there
is no significant difference between the performance in MLAT of
students in Social Studies and those in Comparative Studies at
Essex, and the mean score is significantly higher than the one

19

Histogram A   MLAT OVERALL SCORES

Number
of
Students

Score

Histogram B   MLAT I

21
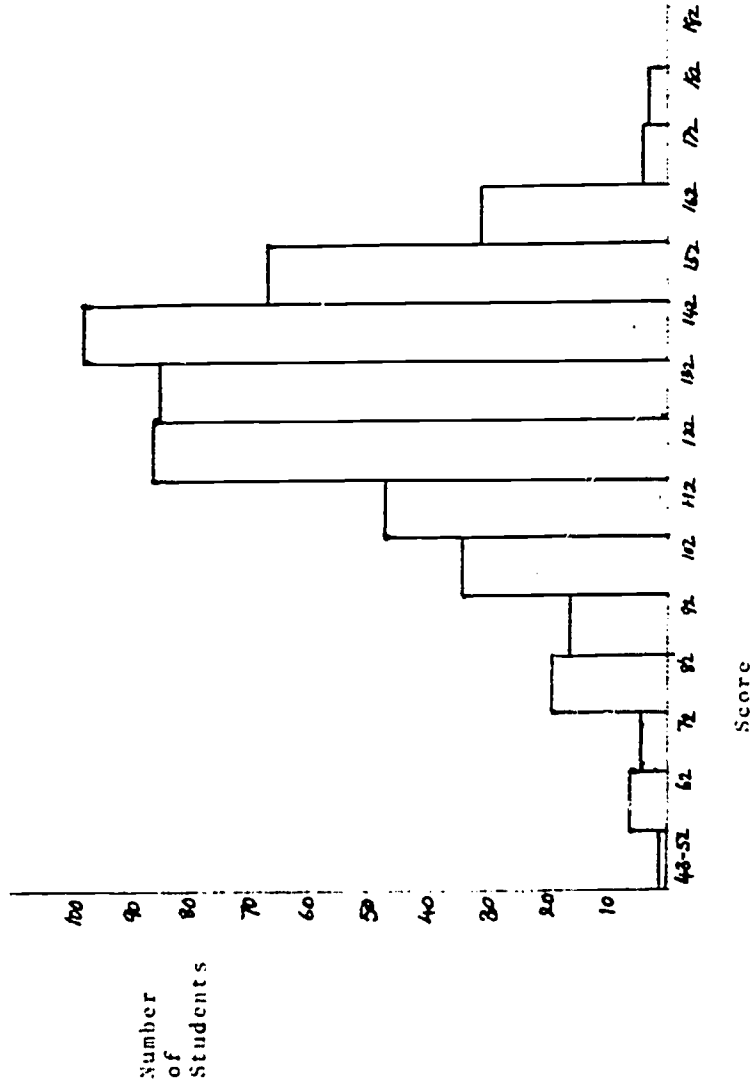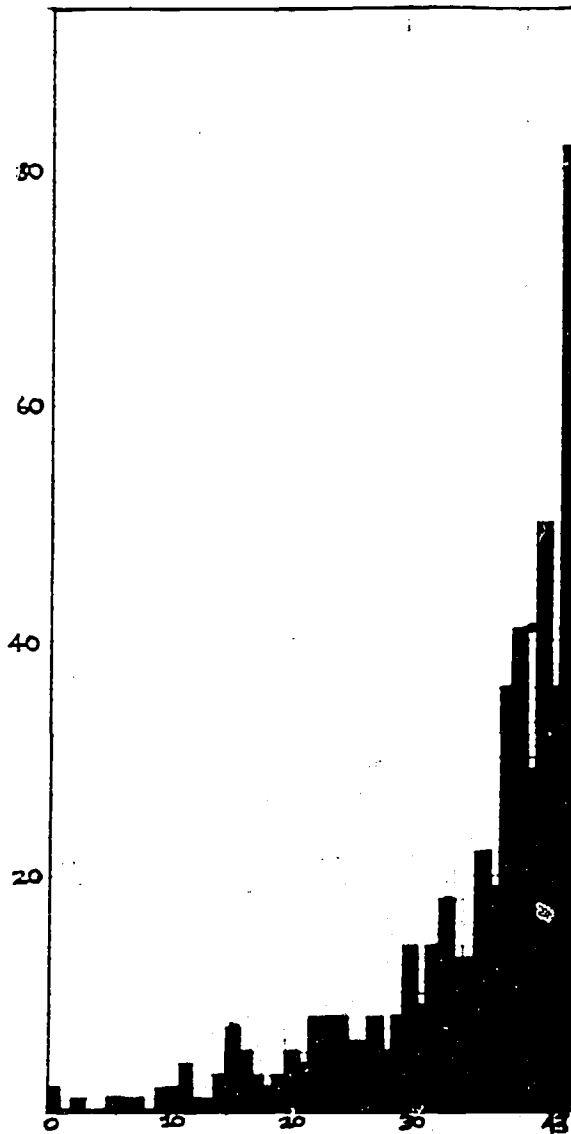
Histogram C   MLAT II

22

16
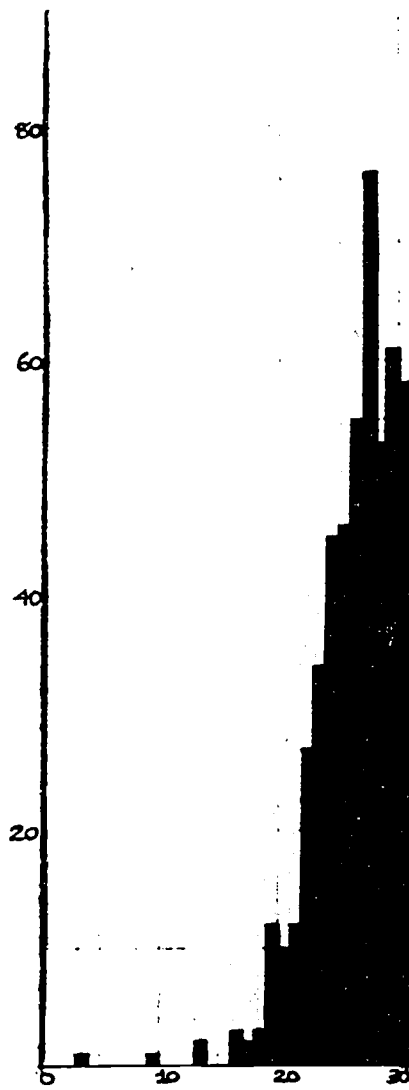
Histogram D   MLAT III
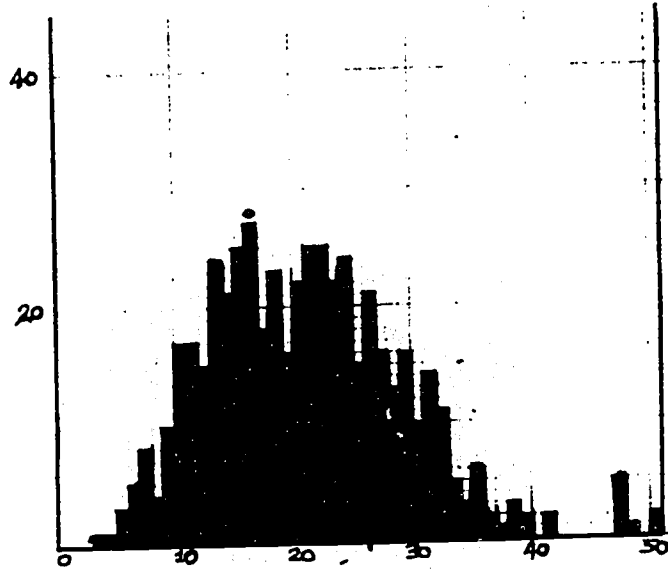
Histogram E   MLAT IV

23
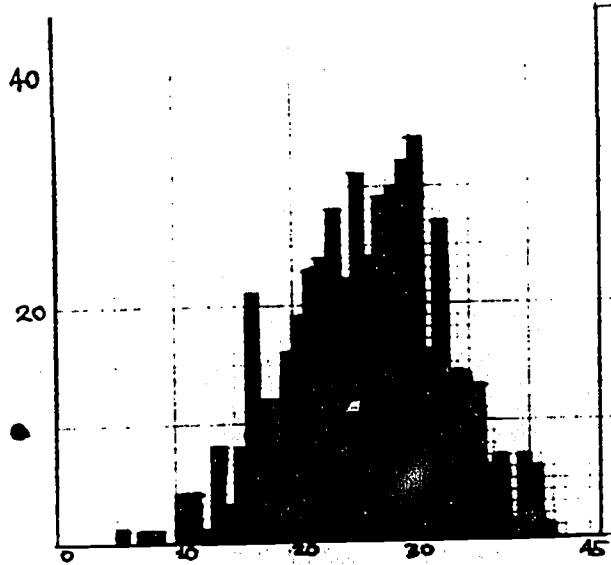
Histogram F   MLAT V
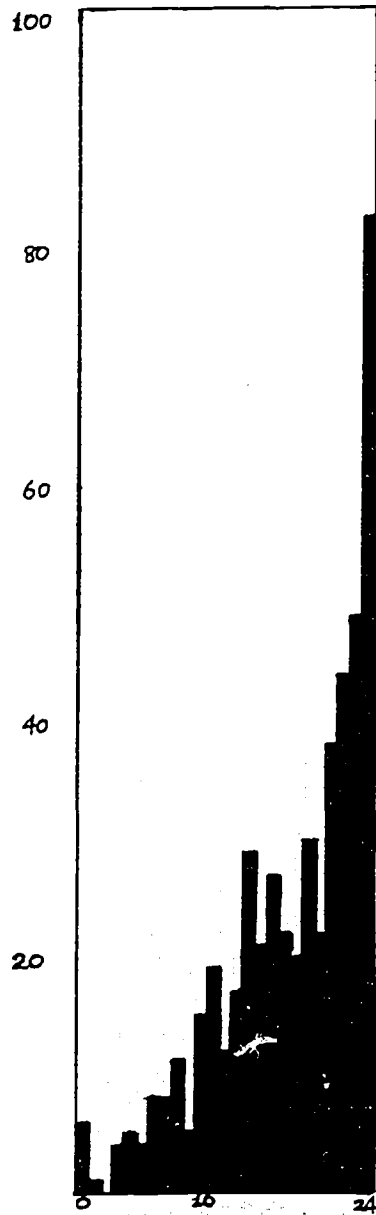
24

obtained by Bath scientists.

The Bath histograms are very similar in shape to the Essex ones, and it is probably not a useful exercise to duplicate them here. The two histograms for Test II do not differ appreciably. A comparative look at the mean scores of the subtests illustrates an interesting point.

Table 5   Comparison of Mean Scores in Bath and Essex.

	Essex	Bath	Difference
MLAT I	34.88	34.0	.88
II	25.81	24.4	1.4
III	20.58	16.0	4.58
IV	25.56	20.9	4.7
V	17.57	15.0	2.57
Total (Full Test)	124.42	110.3	14.13
(Short Form)	63.72	51.9	11.82

By far the greatest difference in the mean scores occurs in Tests III and IV. Tests I, II and V account for only 4.85 (about 5% of the total possible in these tests) of the difference of 14.13. The total difference 14.13 is 7% of the total possible score in the long form of the test. The difference in the short form is 11.82 which is 10% of the total possible in the short form. The short form on its own, therefore, gives a better discrimination between the two groups of students.

Correlations

Table 6

Correlations	MLAT I	MLAT II	MLAT III	MLAT IV
MLAT II	0.425			
MLAT III	0.211	0.295		
MLAT IV	0.417	0.418	0.280	
MLAT V	0.385	0.351	0.232	0.375

From the above it may be seen that Test III has the lowest
correlation figures. The reason for this may be that this test
is the only one to be highly speeded (Iver Shepherd - see below).
The fact of a low correlation may be a positive rather than a
negative factor in relation to the test as it means that it is
doing something which is specific to itself.

Principal Components

Table 7

Factor	MLAT I	MLAT II	MLAT III	MLAT IV	MLAT V	Variance
1	0·621	0·234	0·444	0·467	0·380	85.39
2	−0·146	(−0·029)	0·876	(−0·099)	(−0·153)	22.11
3	−0·373	(−0·148)	(−0·050)	(−0·120)	0·906	5.39
4	−0·479	(+0·014)	(−0·162)	0·858	(−0·090)	4.02
5	(−0·215)	0·980	(−0·087)	(−0·148)	(+0·044)	1.02
						117.93

Factor 1 is probably Carroll's Factor C or 'Associative Memory'.
It would be a mistake to attempt to ally any of the remainder
with Carroll's Factors since each corresponds fairly closely with
a single subtest, as follows:-

Factor 2	MLAT III	English Vocabulary
Factor 3	MLAT V	Rote Memory (Literal)
Factor 4	MLAT IV	Grammatical Sensitivity
Factor 5	MLAT II	Sound-Symbol Association Ability.

The figure in the variance column measures how much variation can
be accounted for by factors e.g. 85/117.93 (approx. 72%) of the
variation between individual students can be accounted for by
Factor 1.

Table 8 Differences between sexes and schools.

	Total MLAT	Comark	Exmark
Women	+2·95	+0·63%	+0·93%
Comparative Studies	+1·16	-0·50%	+0·33%

None of the above figures are statistically significant.

Conclusions

1. It would appear that the test overall is too easy for our students.

2. There appears to be a lack of balance in the test i.e. Parts III and IV are appreciably harder than the other parts and account for more of the internal variance.

3. Test III does not correlate highly with the other tests.

The Predictive Validity of MLAT

P. T. Culhane

All students who enter the School of Compara-
tive Studies have some sort of Language qualifica-
tion. This is not necessarily the case in Social
Studies, but the total number of students in the
sample without any knowledge of a foreign language
[12] indicates that nearly all students who enter
the School of Social Studies also have had some
experience of a foreign language, even if this
experience has been gained only on a casual basis.
It ፡ normal under our system to use previous
attainment in examinations as a predictor, and to
admit students to university on the basis of their
previous success in examinations. The quality of
'O' level and 'A' level examinations as predictors
has often been questioned. In this University the
School of Social Studies keeps a constant watch on
the correlation between 'O' and 'A' level results
and university exams. The most recent observations
(by Michael Parkin and Giles Homewood, 1969 mimeo-
graph) seem to suggest a positive relationship
between the grades of 'O' level and final degree
results in Social Studies. They found ([16] page
10) that 'Overall performance at 'O' level,
measured in % good grades gained, is strikingly
associated with Part II performance.' Thus in
their report quality and not range is found to have
the most significant correlation. A further
observation is that 'performance in Mathematics
and Science O-levels is favourably associated with
Part II performance; this is not repeated with any
other category of subject we measured.' Thus some
predictor variables are more efficient than others,
for the purpose of Social Studies. John Heywood
(Lancaster) in a paper on student wastage [17]
discusses the correlation between 'A' level grades

and psychological tests with final performance
(pp. 8 - 12). He quotes an investigation at
Sheffield by Austwick who found that French 'A'
level correlates very highly with final degree
performance in French. This is likely to be true
with other foreign languages as the skills demanded
at 'A' level are still a necessary component in
'finals' performance. A. Davies [18] finds that
'GQ' (I.Q. score and achievement in major school
subjects) is a good predictor of success in foreign
languages, but his conclusions are very tentative.
We have not tried to correlate MLAT with the
reasoning tests used in the University, as we were
not concerned with the possible use of MLAT as a
replacement for 'A' levels. Our problem is a
slightly different one.

The scope of the investigation

     The aim of the investigation is to see how
well MLAT works in our situation. We already
have a certain amount of information about each
student who enters the Language Centre e.g. we know
how many 'O' and 'A' level subjects he has and we
know his previous attainment in language. What we
want to find out is how well MLAT will predict on
its own and how much it improves on predictions we
may make using only the information we have from
the student's previous language study. It might
be possible to work out a rank order of attainment
by allotting points to 'A' level grades and adding
these together for each student. We decided not
to do this but to take the information in its
crudest possible form. We were left with the
following categories:

    1.   Number of 'A' levels in language. n_A

    2.   Number of 'O' levels in language. n_O

    3.   Number of Languages studied on a
         casual basis (including failed
         'O' levels) n_C

     It was not the object of the exercise to see
whether previous language study made any difference
to the score in MLAT, as we found that at this

stage we were involved in a 'circular' argument
i.e. 'do people study language because they have
high aptitude or do they have a high aptitude
because they apply themselves to the study of
language?' In our situation it was impossible to
be conclusive, indeed an experiment with a totally
different type of population (e.g. at the Secondary
School level) would have to be set up in order to
investigate this problem. I quote the following
Mean Scores for our population by categories,
merely out of interest and so that the work in
producing them might not be wasted.

Previous Language Study	No.	Mean
Students with no language	12	86.17
Students with only 'casual' language study	37	102.54
Students with less than three languages	335	119.30
Students with three languages or more	117	149.91
Overall	501	124.42

A further predictor which was taken into
account is the Audiometry Test which was given to
about 50% of the total population. This is a
standard medical test which is used to measure
hearing loss at different frequencies. The results
are presented in the form of a graph for each ear.
Jack Kay, of the Language Centre, who was in charge
of the administering of the test, and I decided to
classify the results subjectively according to an
arbitrarily defined mean. We produced A, B, C and
D classifications, in an attempt to make the test
measurable numerically. Since the audiometry test
is essentially a method of detecting a physical
deficiency or illness, it is not a test of aptitude,
and it would be a mistake to try and evaluate it as
such. If a person has significant hearing loss, it
is necessary to take this into account in teaching

30

him, in other words, to take remedial measures to
prevent the deficiency having too great an effect
on his studies. The test is designed to point to
deficiencies, not aptitudes, and should only be
taken into account where deficiencies occur.
Interesting work on this topic has been done at
the Reginald M. Phillips Research Unit at the
University of Sussex (see [19] and [20]).

For the purposes of evaluation we decided to
take the overall MLAT mark and its component parts.
We wished to find out the following information:-

a) Whether MLAT as a whole is a good predictor.

b) Whether it is better than previous language
study.

c) Whether the component parts of MLAT would
contribute to a prediction of later attain-
ment and by how much.

d) Whether the prediction we could produce is
more applicable to some languages than to
others, and again by how much.

e) Whether there were any significant differ-
ences between sexes and schools.

The Criterion Variables are what we might expect
MLAT to predict.   In this case there are two
criterion variables, course-mark and examination
mark.   These were available for about 95% of the
population.   The other 5% are people who started
courses and for some reason did not stay long
enough to obtain any assessment, or people who were
not assessed for any other reason.

Bias Variables are a) the different courses i.e.
whether a student is first or second year
                    and b) the different languages
studied.
These are two reservations which one must make at
this stage.   These reflect to some extent upon the
quality of the data.   They are as follows:-

1) The Common First Year French course is
streamed, but since no information on this stream-
ing was readily accessible the data has been
analysed as if no such streaming had taken place.
The result of this is to underestimate the predictive
efficiency of MLAT.

31

2) Multiplicity of Languages. Given that much
higher marks tend to be awarded in Russian, for
example, the statistical analysis attributes this
disparity entirely to a difference in marking
standards. This is not to deny the possible super-
iority of Russian students, but merely to accept
the fact that such things may not be determined
objectively.

Correlations

Predictors	Multiple Correlations		Canonical Correlation
	Course Mark	Exam Mark	
Bias + PLS	0·410	0·547	0·548
Bias + MLAT	0·388	0·518	0·519
Bias + PLS + MLAT	0·433	0·586	0·587

It will be observed from the above that the
multiple correlations are higher with Exam Marks
than with Course marks. In order to assess whether
MLAT is a better predictor than PLS we compare the
figures on the top line with those on the second
line. We would expect the figures on the third
line to be higher since more predictors are
operating. The crucial question is "BY HOW MUCH?"
Thus it is obvious from the above, that

a)   Previous Language Study (PLS) is a
      better predictor of course marks and
      examination marks than is MLAT.

b)   There is an improvement when we add
      MLAT, but this improvement is not very
      significant.

c)   MLAT, when used in isolation, is quite a
      good predictor of examination marks, but
      not of course marks.

We have said [p_2] that the purpose of a
multiple regression is to provide a prediction
equation. Below are listed the prediction equat-
ions for course marks and examination marks. The
parts of MLAT are separated for this purpose.
Where they do not occur in the equation they do

32

not contribute significantly to the prediction.

Course Mark   (x)

__Without MLAT__

$$x = 4.79n_A + 0.88n_0 - 1.17n_C + 48.52 + 9.99 \text{ (if}$$

$$\text{Russian)} + 1.83 \text{ (if 1st year course)}$$

__With MLAT__

$$x = 3.52n_A - 1.45n_C + 0.093m_1 + 0.107m_3$$

$$+ 0.157m_4 - 0.913a + 43.95 + 9.77 \text{ (if}$$

$$\text{Russian)} + 2.11 \text{ (if 1st year course)}$$

From the above it is possible to deduce that
course marks in the Language Centre are on average
well above the 40% pass mark.. It is also possible
to see that not all the component parts of MLAT
have relevance to the prediction of course mark.
The number of languages studied on a casual basis
(n_C) and the mark for the audiometry test (a) have
a slightly negative contribution to the prediction.
We have already decided to disregard audiometry as
a predictor. The figures above confirm that this
is a right decision. The negative contribution of
the number of languages studied on a casual basis
is probably caused by the fact that the figure
includes failed 'O' levels.

Let us take an example and work it out on the
basis of the above prediction equations.

A. Brown has 2 'A' levels and 3 'O' levels in
language. He decides to take Russian in the Prelim-
inary Language Year and his MLAT score is 132 $(36m_1$

$+ 25m_2 + 30m_3 + 21m_4 + 20m_5)$

Mr. Brown's course mark (in round figures) accord-
ing to the prediction figures ought to be

$$\underline{10} + \underline{3} + \underline{49} + \underline{10} + \underline{2} = 72$$

33

If we include MLAT in the prediction we have the following picture,

$$7 + 3.6 + 3 + 4 + 43.95 + 10 + 2 = 71.5$$

The second prediction should be a better one than the first, but the difference between the two predictions is not very high.

Exam marks (y)

Without MLAT

$$y = 7.92n_A + 4.32n_0 + 1.68n_C + 49.41 - 4.74 \text{ (if}$$

French) $- 9.17$ (if Spanish)

With MLAT

$$y = 5.82n_A + 2.74n_0 + 1.05n_C + 0.268m_2 + 0.110m_3$$

$$+ 0.250m_4 + 0.156m_5 + 33.86 - 4.19 \text{ (if}$$

French) $- 8.90$ (if Spanish)

We have already observed that both previous language study and MLAT are better predictors of examination marks than of course marks. This is reflected in the above two equations by the increase in the share of the prediction of 'A' levels in the first one, and the increase in the number of parts of MLAT figuring in the second.

Let us return to Mr. Brown, and see how he would have fared in the prediction. Again in round figures.

Without MLAT

$$16 + 13 + 49.41 = 78$$

With MLAT

$$12 + 5 + 9 + 3 + 5 + 3 + 34 = 71$$

Again, the second prediction is lower and probably has greater accuracy than the first one.

Of course it is not possible to say that the above predictions work in every case. They are

merely suggested by the figures we have obtained.

Conclusions

1.   MLAT is almost as good a predictor of
     likely success in foreign language
     assessment as are previous attainment
     tests in languages, when these are taken
     in their crudest form.

2.   The amount it is possible to add to a
     prediction of likely success when one
     has information about previous language
     study, is insignificant.

3.   As we already have information about
     previous language study in the vast
     majority of cases it would be a waste of
     time to attempt to make this aptitude
     test a universal selection criterion for
     the Language Centre.

4.   The test could be useful in the following
     circumstances

          a)   If a student wishes to find out
     whether or not he has a latent ability,
     which has not yet manifested itself in
     terms of attainment in language learning.*

          b)   In doubtful cases, for the
     'streaming' of students in a large group
     (e.g. French) although it is doubtful
     whether this could be effectively imple-
     mented.

     * In view of the fact that we are not
     certain of whether or not results in
     MLAT have a significant dependence on
     previous language study, do we compare
     the results of a person in this category
     with the overall norm or with people in
     his own category?

Performance and errors in the MLAT of students of
'proved' linguistic ability.

E. Lee

Introduction

In the second term of the M.A. Course in
Applied Linguistics, eighteen of the twenty-three
graduate students of the Language Centre, University
of Essex, sat the full version of the Carroll and
Sapon Modern Language Aptitude Test. (MLAT)

The aim of this paper is to record and, where
possible, evaluate the results of this test session.

A)   First, before looking at the individual sub-
tests and scores, let us run through some of the
basic assumptions implicit in the MLAT.

The MLAT was designed "chiefly to provide an
indication of an individual's probable degree of
success in learning a foreign language" [5:3]*.
Its main use then may be seen to lie in its predict-
ive or prognostic value.   In analysing the results
of the M.A. group, however, we are more concerned
to show whether the MLAT scores reflect the measure
of success one would probably expect of a popula-
tion who, by comparison with a random population,
are of proven linguistic ability.

This type of test predicts how well a person
is likely to learn a foreign language, and is used
to select persons likely to profit from foreign

*   All references for quotations will be shown
thus - that is, number 5 on bibliography, page 3.
Other references will be by author and year only.

language instruction.   [Mackey: 1965]

Since the members of the M.A. group have
already been shown (by examination) to have
profited from foreign language instruction, we may
hope to show the validity for the MLAT.   Its
validity (i.e. how well it measures what it pur-
ports to measure) would usually be calculated by
establishing a correlation coefficient between
test results and the degree of success or failure
in studying, subsequently, a foreign language.
Nevertheless, we should expect to find a strong
positive correlation between 'proven ability' in
foreign language learning and the overall MLAT
scores.   There would perhaps be significant differ-
ences between the F.L. group (those who have
pursued advanced foreign language study - especially
recently) and the English group (those who have
had little or no contact with foreign languages -
in any formal way or for an extended period -
since leaving school).

Further research and continued interest in
the MLAT and similar tests is required since they
fulfil an important role.   The importance of
language courses outside the school classroom is
being recognized at last.   More and more such
courses are provided for people in the services,
in industry and commerce.   This necessarily means
more adults are starting what are often expensive
series of classes.   They are expensive of time to
the employers who release staff to attend Colleges
of Further Education, expensive of staff in those
colleges, expensive of equipment in the form of
language laboratories and so on.   Selection, to
enable the most efficient use to be made of time,
money and resources, is essential.   On what basis
are we to select?   With highly heterogeneous groups
it is not at all a simple matter to choose people
who are likely to benefit from a foreign language
course.   With a group of 'streamed' school children,
where their I.Q. is known, their ability in other
subjects and, more important, their degree of
motivation and application are also known, the job
may be somewhat easier.   Even if all this inform-
ation were available for an adult group, it is not
at all clear that a safe prediction could be made
about success in foreign language learning.   The

37

MLAT presumes to measure abilities (see below)
which are specific to language learning and which
are possessed to varying degrees over and above the
'general intelligence' factor.

It may be argued that all people will learn a
language given enough time and supervision. But
this 'given' is vital. In any real-life situation,
time is probably the most crucial single element.
The MLAT does not attempt to measure whether a
person will learn a foreign language, but how well
he is likely to learn in a given time. The measure-
ment then is relative, rather than absolute.

The MLAT is useful in other ways too. It
possesses a diagnostic function - that is, it can
help to point out specific areas of difficulty in
the language learning task, as revealed by depressed
scores in certain subtests. Furthermore, it may
help to set up 'fast' and 'slow' groups within a
large class. This sort of information is of great
use now that many language laboratories have a
'grouping' facility, allowing different subsets
within the class to work on different material and
at their own speed.

In the writer's school, candidates were selected
for foreign language classes, in which some dubious
hierarchy was agreed upon: Latin, German, Spanish
(French being compulsory), on the basis of their
first year English marks and I.Q. tests. Since
these were taught by grammar-translation methods
which may make appeal more to general intelligence
and problem-solving abilities, there may have been
some justification in choosing prospective Latinists,
etc. in this way. Subsequent results may well have
'justified' the means. It is doubtful whether a
choice based on general abilities would be so
effective for audio-lingual or other orally oriented
courses of study, whereas the MLAT appears to be
equally able to predict success in any type of
course.

Some tests have shown (Harding 1958) that an
initial tryout of a foreign language is a good
predictor of subsequent success in that language.
The MLAT, however, apart from being applicable to
any language, gives the same reliable information
in an hour or so as against a matter of weeks for
the 'tryout' method. On a short-course especially,

the latter would be wasteful of time, yet the MLAT
is often at its most reliable in these circumstances.

Sapon (1955) underlined the danger of relying
on the impact of puzzle-solving ability when speak-
ing of the characteristics of a successful artificial
language for prognostic texts.

"It should present complexities and
resolutions of code structure as reflected
in the development of human language, rather
than in the construction of an abstract code
problem." (p.98)

Carroll (1962) states that:

"Facility in learning to speak and under-
stand a foreign language is a fairly special-
ized talent (or a group of talents), relatively
independent of those traits ordinarily
included under 'intelligence'." (p.89)

(Relatively recent developments in transform-
ational generative grammar too, though for very
different reasons, support this view.) The 'special-
ized' talents are still intellectual abilities and
to some unknown degree trainable [4:89 fn.2]. As
such they are rejected by Pimsleur, Stockwell and
Comrey. (1962).

"It had been hoped that foreign language
achievement could be predicted on the basis of
intellectual factors, such as the ability to
discriminate sounds, to induce grammatical
principles, and so on. Instead, it appears
from these studies that the two biggest
factors in such achievement are the very
general ones of verbal I.Q. and motivation"(p.24)

There would seem to be no obvious reasons why
the MLAT scores should be invalid for the M.A. group
as a whole or for subgroups thereof. The reason is
that the MLAT is largely independent of certain other
variables:

(i) Language

"Fluctuations in the predictive validity
of the MLAT do not seem to be consistently
correlated with language or even type of
language." [5:21] [cf. 2:158].

(ii) Previous language training

"All the evidence which has come in to
date suggests that performance on the test
itself is not much subject to improvement
solely through taking foreign language courses."
[2:157]

Yet,

"Amount of previous language training
and success in previous foreign language
courses (important in our group - E.J. Lee)
probably constitute predictors which may make
some unique contribution to prediction over
and above the MLAT" [5:21]

Any help is likely to be of an indirect nature.
That is, previous experience may give the student
a better idea of his own learning strategy - more
specifically, how to organize, for example, the
learning of the Paired Associates in Part V.

(iii) Language Method

"There was no systematic fluctuation of
validity dependent on teaching methodology"
[5:22]

Finally, the MLAT has been validated for
literate persons of native (or near-native) fluency
in English. All members of the group, it was
established by questionnaire, are native speakers
of English.

The full version of the test which, other
things being equal, is reckoned to be more reliable,
was administered to the group under nearly identical
conditions in the language laboratory. One or two
members who were unused to the laboratory conditions
experienced a certain fatigue but there is no
evidence to show that this affected the results.

Since all persons wore headphones any advantage
gained in tests I and II [see 5:4 fn.3] would be
equally shared.

Following Carroll (1962), we may define
aptitude as: "the time which would be needed by
individual 'i' to learn task 'j' to a specified
criterion of learning, on the assumption that...
the task is presented well enough for him to

understand the task in the light of his general
intelligence." [4:122] [cf. 2:159]

More particularly:

"The MLAT measures certain learned
capabilities of the individual which are
apparently prerequisite to reasonably rapid
success in learning a foreign language." [5:21]

That these are 'learned' presumably implies
that they are learned during L_1 acquisition and
are applicable to L_2.

What these capabilities are has been variously
assessed. Compare with the following, those given
by Mackey (1965) p. 326 and Pimsleur (1964). The
latter hypothesizes that auditory ability is the
factor which accounts for differences in people's
language learning ability which are not explainable
by intelligence or interest.

Carroll (1958) lists seven factors: (see pp.11-16).

Factor A: Verbal knowledge. This comprises a know-
    ledge of the vocabulary and structure of L_1,
    which is not unlike verbal intelligence.

Factor B: Linguistic interest. One of the most
    important yet most difficult to assess, this
    factor is a specific motivation, interest or
    facility with respect to unusual (presumably
    'novel' rather than 'weird' - E. J. Lee) ling-
    uistic materials.

Factor C: Associative Memory. This includes
    immediate (rote) memory.

Factor D: Sound-Symbol Association. This is the
    ability which represents the extent to which
    the individual possesses a knowledge of sound-
    symbol correspondences or can learn a novel s
    set of such correspondences.

Factor E: Inductive Language Learning Ability. The
    ability to induce grammatical rules and
    properties of a language.

Factor F: Grammatical Sensitivity or Syntactical
        Fluency. That is, a sensitivity to the

41

functions of words in sentences and a facility
in producing coherent verbal materials.

Factor G: "<u>Speed of Association</u>".   This appears to
have no significant role in foreign language
learning, and will be ignored hereafter.

Not all these factors are of even weight of
importance. Factors B, C, E are of greater weight
than factors A, D, F.

How far and which of the above are measured in
the MLAT we shall see more precisely in the following
sections, when we look at the different subtests one
at a time.

In the considerations of the various subtests,
we have thought it useful to do an error analysis,
in the hope that this might lead to constructive
criticism of the design of the MLAT, and also to
some real measure of internal validation of the
test - (see item 1g on P.T. Culhane's circular of
22.11.68) - if not immediately, then at some future
date when other data can be considered together with
that provided herein.

B)   TEST I Number Learning.

"This seems to measure one aspect of the
memory component of foreign language aptitude,
but the part also has a fairly large specific
variance, which one might guess to be a special
'auditory alertness' factor which would play
a role in auditory comprehension of a foreign
language." [5:3]

Since this and Part V both measure memory in
some form, it is interesting to note from Fig.1
that there is a very strong similarity between the
shapes of line I and V on the graph. Note the
lower score on both tests for Nos. 7, 8 and again
for Nos. 11- 12, 15 -16.

Speed was no problem here; all members of the
group completed the test and the marks were high.
A mean score of 38.3 out of 43 (or 89.1%) meant
that this test had the second highest mean score.

42

Error Analysis:

TOTAL: 69 errors

The number items to be learned were composed
of English sound elements. An analysis of the
answers required, in the main, hundreds, tens and
units to be recorded. With the HUNDREDS there were
13 errors, with the TENS there were 33 errors, with
the UNITS there were 23 errors. This would seem
to suggest that in being presented with an unfamiliar
sequence XYZ, then attention and therefore retention
is better in the case of X, Z - that is the first
and last elements in each case. Y, the middle
component, is less well perceived and/or recalled.

With regard to the test design, the process of
recopying the digits into the marking slots appears
to present problems.

3 people miscopied 1 item
1 person miscopied 5 items
1 person miscopied 7 items

Of the 15 items miscopied, 13 were failures
to mark zero on the score sheets. The trouble
appears to arise since there is no zero mark in
the hundreds column, but only in the tens and units
column.

The case of zero occurs in another context too.
Ten of the 69 erros were on items c., n. - that is,
with no 'hundreds' digit required in the answer.
Failure to distinguish 32 from 302 (item n) costs
the subject one mark - the units are correct, the
tens wrong (lose one mark) and the hundreds ignored.
But, to write 34 for 304 (item d for example) costs
the subject 2 marks - one for the tens and another
for the hundreds. This would appear to be a strange
inconsistency which might be removed.

The errors for the various items break down as
follows:
(a) 9 (b) 9 (c) 8 (d) 1 (e) 4 (f) 4
(g) 5 (h) 6 (i) 4 (j) 0 (k) 3 (l) 1 (m) 7
(n) 8 (o) 0

The high error rate at the beginning may be
due to hesitation in the early stages of this first
test.

Of the 69 errors, 15 were miscopyings - see above. The remaining 54 were comprised thus:

    (i)   34 confusions (e.g. 200/100 or 0/40).
   (ii)   13 wrong column (i.e. 10 confused with 100
                or 4 with 40)
  (iii)   7 omitted (i.e. 242 written as 202).

<u>Note</u>:

    (i)   31/34 of the confusions were between numbers of the same power of 10 - i.e. 100/400: 1/3. Only 3 were of the type 2/40 - that is a double confusion between the basic root number 2/4 <u>and</u> the column units/tens.

    Further <u>26</u> were confusions between consecutive digits 1/2, 3/4, 2/3, whereas only <u>6</u> were between 1/3, 2/4 and only <u>2</u> were between 1, 4.

    This would indicate that the closer the digits were together the more readily were they confused.

   (ii)   The 13 'wrong column' errors were well spread over the 4 digit root-words (1-4).

          2 wrote 40 as 4
          3 wrote 10 as 1
          4 wrote 20 as 200
          4 wrote 30 as 300

   N.B. The tens column is at the bottom of the confusion.

   (iii)   Since there was no number to be learned for zero, it is strange that 7 should have written in unnecessary zeros.

    There appears to be no pattern of any one subject consistently making errors with any one number or digit.

   C)   <u>TEST II</u> <u>Phonetic Script</u>.

    "This appears to measure what we have called sound-symbol association ability, that is, the ability to learn correspondences between speech sounds and orthographic symbols. It may also measure a sort of memory for speech sounds, and it tends to correlate highly with

the ability to mimic speech sounds and sound
combinations in foreign languages." [5:3]

The three 'mimics' of the group (subjective
impression) each scored 100%, but then so did most
people. Speed, once again, was no problem. Every-
one finished the test, 10 out of 18 scored 100%.
On the face of it the test was too easy. It was
the best done of the five subtests, the mean score
being 29.2/30 or 97.4%.

Can one argue that the test is too simple?
Certainly one would have expected the scores to be
high from our group. First, the sounds were those
of English and thus familiar. Second, the sounds
were distinctive or phonemic; there were no slight
allophonic variations to discriminate. Third, every
member of our group has some background in phonetics
and at least a passing knowledge of the phonetic
script, although the diacritics on š, ž, ʒ, etc.
are American rather than British. Fourth, every
test item was of a CVC arrangement, allowing maxi-
mum distinction between sequents. Consonantal
clusters may have been harder.

Error Analysis:

14 items were wrong, of which 13 were minimal
pair errors and 1 a multiple error.

The items wrong were:

   Once each: 6, 11, 12, 14, 18, 20, 24, 30.
   Three times each(!): 26, 27.

(i) Five people were wrong with the ə/ô
distinction. It seems unlikely that this caused a
problem in discrimination, and seems more probable
that the errors were due to the fact that these
were the most 'exotic' of the symbols used:

         Symbols:

Alphabetic: t k ʒ ɑ ɛ ʒ   i ə e
Exotic: š ž ʒ č ɔ ô   iy aw ey ay ə æ

Two more subjects hesitated on ô/ɔ before
getting the correct answer by changing their minds.

(ii) Two errors were made on the length distinc-
tion in i/iy.

(iii) <u>Three</u> errors were made on ay/iy, which I
would presume to be a confusion of orthography and
symbol. On hearing /ay/. the sight of 'iy' on the
page may appear the right version by association
with the sound of these letters in normal ortho-
graphy.

(iv) The remaining errors were a/e; æ/ə; j/ʒ
and ə/ey.

Note that the consonant/vowel errors are approx-
imately even:- 6:8 respectively.

One which this writer had trouble in hearing
was the difference between the American vowel /æ/
'<u>bag</u>' and /a/ '<u>dog</u>'. Surprisingly, no errors were
made in items containing this opposition, though
there were two occurrences of hesitation - that is
marking the wrong one first and then changing one's
mind.

D)   TEST III   Spelling Clues.

        "Scores on this part depend to some extent
    on the student's English vocabulary knowledge.
    This subtest also measures the same kind of
    sound-symbol association ability as measured
    by Part II, Phonetic Script, but to a lesser
    extent.   It is highly speeded." [5:3]

There is no evidence from our scores that the
English group did any better on this test than the
foreign language group.

Parts II/III may be thought of as measuring
differing aspects of the same ability, Part II to
hear phonemic distinctions and Part III to produce
phonemes. [cf. - 4:115]. With the small numbers
available, the scores for these two tests do not
seem to correlate to any extent. (See Fig. 1).
Certain members of the group felt there may be a
negative correlation since the 'rigour' of the one-
sound-one-symbol presentation of II may have caused
trouble when coming to Part III. However this seems
unlikely, since the 'spellings' in III are in fact
quite 'regular', once the pattern has been learned.

The test certainly was speeded - (see details

below). This test was the worst done, the mean
score being 27.7 out of 50 or 55.5%.

Error Analysis:

There were 60 errors altogether - that is 60
items marked wrongly. These are shown in Table 4,
listed in order of overall rank order i.e. accord-
ing to total raw scores). Most tables have been
arranged in this same order to allow cross-compari-
son.

There appears to be no connection between the
number attempted and the number correctly done.
The 'number done' is that number from the start to
the last one attempted.

A complete survey of the errors analysed in
terms of the multiple choice items is given in
Table 6. Certain of these - bracketed thus: [ ]
on Table 6 - require comment here.

Firstly though, item 13 'frajl' (fragile)
would seem to be more easily recognized by an
American than a British speaker. However, this
item caused no error.

One general comment would be that the method
of answering, no doubt because it was a speeded
test, seemed to be to mark the first possible
correct multiple-choice item and move on, rather
than reading all 5 possibilities and selecting one.
This judgement is based on the observation that
only 6 positive errors occur after the correct item
(marked X).

If all the multiple choice items are contribut-
ing to the test equally as distractors (if not them-
selves the correct item) then one would expect the
errors to be evenly spread, yet this is not so in
every case.

(i) Item 2. 'rgument'. All 5 errors were on
A choice - 'regiment'. The closeness of the spelling
has overriden no similarity in their meaning.

(ii) Item 4. 'nme'. 6 of the errors were on
A choice, 'sea-animal', presumably because of the
sound-association with 'anemone' (?)

(iii) Item 8. 'ple' (plea). 4 people omitted this - more than any other item. The meaning link with 'appeal' (choice D), is perhaps a little remote.

(iv) Item 12. 'ndkat'. 3 out of 5 chose C - a 'species of feline'. No doubt the last 3 letter of the item are of significance!

(v) Item 14. 'snser'. 3 out of 5 chose C - 'respond'. One can only guess at a connection, say with 'sneer'?

(vi) Item 22. 'kurs'. 6 out of 6 chose D - a 'race-track' - representing a confusion of 'course/ curse'.

(vii) Item 50. Only one person attempted this, and he got it wrong. The item 'pes' followed by c choice A, 'vegetable' is likely to read 'peas' with a voiced /s/ finally. However, the final /s/, consistently with other items (cf. 6,22) was voice-less.

E)   TEST IV   Words in Sentences

"This part is thought to measure sensitiv-ity to grammatical structure, and may be expected to have particular relevance to the student's ability to handle the grammatical aspects of a foreign language. As yet, it is not known how much scores on this part are a reflection of formal training in grammar; at any rate, no grammatical terminology is involved, so that the scores do not depend upon specific memory for grammatical terminology." [5:3]

Whether scores on this test reflect formal training or intuitive knowledge is not known. With the amount of formal training done by this group one might have expected the scores for this test to be higher. Most members of the group found it difficult to assess how well or badly they had scored on the test, yet most felt this was their worst score. In fact this was not the case for the majority of sub-jects. A mean score of 31.4 out of 45 or 69.9% rakes this test the fourth best done overall.

When comparing it with III (see tables 4 and

3) we note that there were more errors made on this
test (195 to 60) on fewer items (45 against 50)
and yet only three people did not complete, or
nearly complete Part IV. After test III, which was
highly speeded, some degree of learning may have
ensued, which prompted subjects to work far more
quickly in Part IV - even at the expense of accur-
acy.

On Figure 1, we see that the curve for Part IV
gives the closest approximation to the overall-
score curve. The short form of the MLAT includes
Part IV, indeed perhaps this is the best single
test to record aptitude for language learning.

Error Analysis:

The complete breakdown of errors is given in
Table 7. Once again, as in Part III, certain items
call for immediate comment. The increase of errors
towards the end may be due to:

(a) increasing difficulty of the items - though
we can find no reference to this, nor is it detect-
able from a rescrutiny of the items;

(b) fatigue of the subjects;

(c) increasing speed of work before the end of
the time limit;

(d) any combination of the above.

Without counting those errors-by-omission
listed in Table 7, there are still a great number
of errors, not all of which are evenly spread over
the remaining four incorrect items available for
each item.

(i) Item 9. That two people should omit this
at an early stage in the test seemed significant;
the stimulus item 'SPEEDING', too. On investigation,
items 9,20,32,33,42,43 also contain -ING stimuli
and count among the highest error items in the test.
In each case there are confusions between Nouns,
Adjectives, Participles and Gerunds in -ING.

Item No. 7 is, in this sense, exceptional.

(ii) Item 11. Comments on item 9 show that
general similarity of spelling, part of speech and
so on seemed of more importance in determining which

49

item was chosen than <u>function</u> in the sentence.
Item 11 too shows this. Bill (S) and (*R) Grant
are both proper nouns.

(iii) Item 12. Again, surface similarity occurs.
<u>Piece</u> of N is equated with <u>dawn</u> of N in spite of the
fact that the first is Obj. N. and the second is
part of a proposed ADV.PHR.

(iv) Item 16. Again, compare '<u>swollen</u>' (s)
and '<u>hidden</u>' (wrong Res).

(v) Item 26. Conjunctions - badly done on the
whole. Compare items 1, 15, 34.

(vi) Items 31/40. Pronoun is linked with pro-
noun.

(vii) Item 36. Yet again, the two quantifiers
are linked erroneously - by 9 people (where N=18!).
The meaning too could play a part, since 'many' and
'a number of' may be synonymous.

(viii) Items 37/45. These two demonstrate clearly
the surface connection we are claiming, which has
been the greatest single factor in causing error.
Item 37, especially shows that choice D - the only
choice containing -SELF was wrongly selected by 10
subjects.

(ix) Items 38/44. Permit of no 'simple'
explanation.

(x) Item 41. '<u>it</u>' object of the verb '<u>take</u>'
is linked with '<u>steps</u>' perhaps also thought <u>to be</u>
the object of the same verb.

F) **TEST V**. Paired Associates

"This part measures the rote memory aspect
of the learning of foreign languages." [5:3]

All completed this test in the time allowed.
The mean score was 17.5 out of 24 or 74.1%, making
it the third best done subtest of the five.

At the time of taking the questionnaire answers,

all subjects were retested on this subtest to see
if any items could still be recalled. Listing
subjects by their overall rank-order numbers,
No.1 scored 4; No. 3 - 1; No.4 - 2; No.14 - 1;
No.18 - 1; all others - zero. This would seem to
indicate that the test measures short, rather than
long-term memory.

The writer was further interested to know
whether the possession of a conscious learning
strategy improved performance on this test. From
the answers to the questionnaires we gather that the
eight subjects having such a strategy were placed
in this subtest (not overall scores) as follows:

(a) 4 scored 100% (1st =)

(b) the remainder were 5th, 6th, 7th, 9th
respectively. (N = 18).

On such small numbers even, the evidence seems
conclusive.

Error Analysis:

The complete analysis is shown on Table 8.
Three brief comments are required:

(i) In item 1, 4 people chose A in spite of
there being no word for 'in' included in the list
of associates to be learned.

(ii) Of the other errors, none seems to be of
immediate significance.

(iii) Some of the more 'obvious' distractors
did not cause errors. For example:

No.12: 'hon ≠ on';   No.14:   'ja ≠ yes'
No.17: 'wener ≠ never';   No.21:   'kete ≠ kite'

Only one person wrongly chose 'young' for 'yong'.
(No.18).

G)   A few additional comments are necessary.

1.   It is clear from the lack of homogeneity
of the curves in Figure 1 that the various subtests
of the MLAT appear to be testing quite different,
that is, independently measurable components of
aptitude. Compare Carroll (1959):

"The separate parts of the MLAT are not

highly intercorrelated, and it may be assumed
that they measure somewhat different abilities"
(p. 159).

and this can be seen to be a good thing, for:

"one critical consideration in this
connection is that if two parts are highly
intercorrelated, one or the other may be
failing to provide unique information." [5:17]

2.   Owing to the comparatively small size of
the group and overall homogeneity there is little
evidence to suggest that language aptitude changes
with age: [cf. 5:23]. Also [2:15]:

"It appears that the test measures func-
tions which do not change greatly from adoles-
cence to adulthood."

The overall rank-order positions tabulated
with the age-groups can be seen below.

Table 9.

Age	Positions
25-30	1, 5, 14, 16.
30-35	2, 4, 6, 10, 11, 12, 15, 18.
35-40	7.
40-45	3, 9, 17.
45-50	8.
50-55	13.

3.   With regard to the relative scores of the
two sexes, it has been noted [5:23] that in the
upper grade levels the girls score higher. The
statement that there is too little data available
on adult women is echoed in the present case, only
one of the 18 subjects being a woman.

4.   When dividing the M.A. Group into 'English
only' and Foreign language sub-groups, we find that
once again no clear pattern emerges as to their
relative success in scoring high. If anything, the
results are what one would predict - that is, four
of the six English group are well down below half-

52

way in the overall order.

The two groups are as follows:

Table 10.

Group	Positions (overall)
English	3, 6, 12, 15, 16, 18.
Foreign Language	1, 2, 4, 5, 7, 8, 9, 10, 11, 13, 14, 17.

5.   It has been said [5:23] that the MLAT
shows higher validities for predicting success in
'intensive' courses, where the 'language bath' or
maximum exposure technique is likely to be followed
and where motivation is higher at the start of the
course and likely to be maintained throughout, owing
to the relatively short duration of the course in
question.

The MLAT itself, however, does not pretend to
measure motivation itself. The writer included on
his questionnaire certain questions in an attempt
to obtain some rough guide to the subjects' attitudes
to the test from which it may have been possible to
guess at the degree of motivation.

From the results of the questionnaire (see
paragraph H below) it would seem that even this
rough guide does correlate with final raw-score
on the overall test.

6.   Although the scores at the top end are
fairly close together, the overall range of marks
is not as homogeneous as one may have thought. The
variance does not seem to be attributable to any one
factor. The most interesting (and perhaps signif-
icant) reflection of the scores is in the 'motiva-
tion scale' (see below).

H)   Finally, a word about the questionnaire and
the data collected. (See also Appendix A, b).

The questionnaire was passed round approximately

53

ten weeks after the MLAT was given.

(The numbers of the paragraphs correspond with the question numbers of the questionnaire - See Appendix A.)

1. N = 18: 17 men; 1 woman.
3. All have English as mother tongue.
2,4. These questions were to give information about the language background of the subjects. The range of languages was considerable. No one had studied less than 3 foreign languages and no one more than 8. The levels attained ranged from 'casual' to honours degree standard. The languages studied are/were: Arabic: Cambodian: Chinese (Case-study): Fijian: French: German: Greek (Ancient): Greek (Modern): Irish: Italian: Latin: Malay: Polish: Russian: Spanish: Urdu. (16 in all). All formal training in these languages had been by the grammar-translation method.
6. Degree subjects: (with regard to overall positions):

Table 11.

Sociology	- 18.
English	- 3, 6, 12, 15, 16.
English & Foreign Language	- 9, 13.
Modern Languages	- 1, 4, 5, 7, 8, 10, 11, 14, 17.
Classics	- 2.

5. Languages taught: English; English as L_2/foreign language; French; German; Irish; Italian; Latin; Russian:
7,8. The aim was to establish a 'scale of motivation' thus:

From strongest to weakest: 8c: 8a: 8b: 8d. (with additional material from q.7 if any).

9. Similar to 7/8 with the scale running from 9a-9e.

As was stated above, (Paragraph G) the replies to these questions do bear a relation to the scores achieved.

· 54

Table 12.

Rank Order	7/8	9	10
1	c	a	c
2	a	a	c
3	c	a	expected to do well*
4	d	c	b
5	c	a	b
6	?	?	a
7	b	b	no reaction
8	c	a	
9	?	?	c
10	b	c	c
11	a	c	no reaction
12	b	a	a
13	b	defeatist*	a
14	d	b (+panic)*	a
15	d	b	c
16	d	b	a
17	b	defeatist*	a
18	a	c	b

* These were actual answers given. We thought it revealing to include them.

10. See Table 12 (above).

We felt that people who had expected to reach a high grade might have done better just for that reason. Note that the first three obvious-ly expected to do well. For those who answered (10c) and were lower down the scale, the 'I told you so' takes on quite another meaning.

It is interesting to note that of the five members of the M.A. group who did not take the MLAT, two were absent and the other three said they were not interested in their scores nor in the test - mainly because they were 'English' and the test was concerned with aptitude for foreign language learning.

11. If people consciously used a learning strategy in test V, then the choice was to use mnemonics of one kind or another.

Those placed 1st - 7th, 9th did have some strat-egy.
Those placed 13th, 15th made some attempt.
The rest (8th- 10th, 11th, 12th, 14th, 16th, 17th, 18th) made no conscious effort to employ a strategy.

The figures appear to be significant.
13. To enable the subjects to recall mnemonics used,
    they were shown the list of paired English-
    Kurdish words. Therefore, this necessarily had
    to follow q.12 - the retest. (See Appendix B).

CONCLUSION

    Unfortunately, significant generalizations are
not possible from such a small number of results.
It is hoped that the results of this group will be
useful in research with other M.A. or similar groups,
which may show up special features about results at
the top end of the scale. This information may be
useful to University staff organizing language
courses - especially for non-specialists.

56

Copy of Questionnaire

1.   Name.

2.   Are you doing a case study? Russian/ Chinese/None.

3.   Mother tongue. Specify if __not__ English.

4.   Foreign language(s) studied:

        Language/Length of /How long since/Method*
               Study   last studied

S.   Language(s) taught, (including English) and for
     how long.

6.   Degree subject(s).

7.   Why did you take the MLAT?

8.   Were you interested in (a) your own performance
     (b) the test itself (c) both (d) neither? (that is
     no special reason).

9.   How would you assess your work and attitude whilst
     doing the test (a) all-out effort (b) keen to do
     well (c) take what comes (d) not really trying
     (e) utter waste of time.

10.   With regard to your own score, were you (a) dis-
      appointed (b) pleasantly surprised (c) 'proved'
      to be right about your own attitude?

11.   Did you use a learning strategy in Part V - the
      Paired Associates Test? Specify.

12.   (RETEST of Kurdish vocabulary)

13.   Can you remember any of your mnemonics, etc.?

14.   Any additional comments.

      *(a) bilingual/in the home (b) audio-visual (c) 'oral'
      method (d) direct method (no English at all) (e)
      Grammar-translation (f) casually/short visits to
      the country concerned (g) other (please specify).

1.   Two strategies which failed were:

        (a) on the basis of previous sub-tests, patterns
     of a morphological kind were looked for in the Kurd-
     ish words to separate Nouns from Verbs, etc. This
     wated precious time.

        (b) Links between Kurdish and Arabic (known to
     the Subject) were looked for - in vain.

2.   Three other more helpful strategies were:

        (a) To do the ones learned and guess the rest.

        (b) Remember items by their position on the
     page relative to other items.

        (c) Find the short words - for example 'ja',
     'e' were the easiest to learn.

3.   Other mnemonics were as follows. We have divided
     them up into sub-groups which should be self-explan-
     atory. In each case, the words underlined are the
     Kurdish and English equivalents, the other words
     being the link word or words.

     A.   Sound Associations

            ja - /dzei/ - /dei/ - day.
           *hui - whee!! - fall.
           *kete - camel (alliteration).
           *chomco - Chomsky - body.
            hon - shuddering noise - cold.
            yong - young(hawk) -hawk.
            lah - open-mouthed sound - (open-mouthed)
                                                wolf.
            xozo - so-so - casual - easy.
        cf (nente - dainty sound - lady.
           (nente - feminine sound - lady.
        cf (nente - Italian sound - Italians - lady.
           (nente - dolce far niente - Italian - lady.
            ngoz - a dark word to say.
            mi - 'me' touch me - touch.
            hui - 'high' - high fall - fall.

     B.   The look of the word

           *tsep - step - step inside - inside
        cf (lohong - holong - how long - question - ask.
           (lohong - long for - ask for - ask.
            e - like a bowl on its side.

        Note. One French speaker had to fight off a
     negative transfer for 'jate = sun'
            jate - jatte - (Fr.) bowl.

C. <u>Meaning</u>

<u>roo</u> - kangaroo - the bush - primitive -
cave paintings - <u>art</u>.
<u>wener</u> - Vienna - '<u>Kultur</u>' - <u>book</u>.
<u>kete</u> - Katie (personification of) - <u>camel</u>.

D. <u>Connections with other languages</u>

<u>lohong</u> - like a Malay word - the "Malay
word" means to <u>ask</u>.
<u>xozo</u> - like a <u>Mexican</u> Indian word.
<u>hui</u>/e/mi/<u>kete</u> - like Malayo-Polynesian
words.

Note the coincidences of similarity marked by:

* = identically used by more than one
person.
cf = similar.

Table 1

RAW SCORES - listed in rank order of totals and
equivalent %iles from 'College Freshmen' scale.

Rank Order	I	II	III	IV	V	Total (raw)	%ile College Freshmen
1	43	30	38	36	24	171	99
2	40	30	34	42	24	170	99
3	42	30	35	38	24	169	99
4	40	30	28	40	24	162	99
5	38	30	38	32	23	161	97
6	35	30	28	38	24	155	97
7	36	30	38	33	16	153	95
8	39	29	42	28	15	153	95
9	43	30	24	29	24	150	95
10	42	30	22	33	22	149	90
11	41	28	27	29	20	145	85
12	36	29	31	34	13	143	85
13	40	29	26	30	12	137	75
14	41	30	14	29	18	132	70
15	32	29	24	31	13	129	65
16	31	29	26	28	4	118	50
17	30	29	12	24	10	105	35
18	41	24	13	12	6	96	20
Total Poss.	43	30	50	45	24	192	-
Mean	38·3	29·2	27·7	31·4	17·5	144·3	-
S.D.	-	-	-	-	-	20·95	-

Table 2

MEAN SCORES OF SUBTESTS AND THEIR ORDER

Part	I	II	III	IV	V
Mean Score	38·3	29·2	27·7	31·4	17·5
Mean Score (as %ages)	89·1	97·4	55·5	69·9	73·1
Order of Success - based on %age mean scores	2	1	5	4	3

60

'Table 3

RANGE AND MEDIAN SCORES FOR SUBTESTS AND TOTAL

	I	II	III	IV	V	Total
Range	43-30 (13)	30-24 (6)	42-12 (30)	42-12 (30)	24- 4 (20)	171-96 (75)
Median Score	36·5	27·0	27·0	27·0	14·0	133·5

Table 4

SUBTEST III

Rank Order Overall	Done	Score	Wrong
1	39	38	1
2	37	34	3
3	37	35	2
4	37	28	9
5	38	38	0
6	29	28	1
7	42	38	4
8	50	42	8
9	25	24	1
10	24	22	2
11	37	27	10
12	36	31	5
13	29	26	3
14	15	14	1
15	26	24	2
16	28	26	2
17	13	12	1
18	18	13	5
TOTAL			60

Table 5

SUBTEST IV

Rank Order Overall	Done	Score	Wrong
1	45	36	9
2	45	42	3
3	45	38	7
4	45	40	5
5	45	32	13
6	45	38	7
7	45	33	12
8	45	28	17
9	45	29	16
10	45	33	12
11	43	29	14
12	45	34	11
13	43	30	13
14	36	29	7
15	45	31	14
16	43	28	15
17	27	24	3
18	39	12	17
TOTAL			195

Table 6

ERROR ANALYSIS OF MULTIPLE CHOICE
ANSWERS IN SUBTEST III.

No. Item	Choices						Total	U
	A	B	C	D	E	0		
1			1	X		1	2	-
2	[5]				X		5	-
3	1	X					1	-
4	[6]		X			1	7	-
5		1		X			1	-
6	X						0	-
7	1	1		X		2	4	-
8			1	X	1	[4]	6	-
9			X				0	-
10	1				X		1	-
11			1	X			1	-
12		X	[3]			2	5	-
13			X				0	-
14	[3]	X				2	5	1
15	1		X			2	3	1
16	1	X					1	2
17		X				1	1	2
18				X		1	1	2
19	X						0	2
20				X			0	2
21		X					0	2
22				[6]	X		6	2
23		X	1				1	2
24				X			0	3
25			X				0	4
26			1	X			1	5
27					X	1	1	5

62

No. Item	Choices						Total	U
	A	B	C	D	E	O		
28		X					0	6
29	X						0	8
30			X			1	1	8
31		X				1	1	8
52				X		2	2	8
33			X				0	8
34			X				0	8
35					X	1	1	8
36			X				0	9
37			X				0	13
38		X	1				1	14
39			X				0	15
40					X		0	15
41		X					0	15
42			X				0	16
43			X				0	16
44				X			0	16
45			X				0	16
46		X					0	16
47		X					0	16
48				X			0	16
49			X				0	16
50	[1]	X					1	17

KEY: A,B,C,D,E = Multiple-choice possibilities.
     No. Item = Number of item in test.
     O   = Error by omission
     U   = Unattempted. No. of those not
                 completing the test so far.
     X   = The correct answer.
     1 etc.   = Number of people choosing this
                 wrong item.
     [ ]   = Mentioned in text as significant.

Table 7

ERROR ANALYSIS OF MULTIPLE-CHOICE
ANSWERS IN SUBTEST IV.

No. Item	Choices						Total
	A	B	C	D	E	O	
1		1	X			1	2
2		1			X		1
3	1		X				1
4				X			0
5				X	1		1
6		1	X				1
7	X						0
8		1	X				1
9			X	1		[2]	3
10	2	X		1			3
11				[4]	X		4
12	[5]		X	2			[7]
13		2	X				2
14		1		1	X		2
15	3	2		1	X		[6]
16	X			[4]		1	5
17		1	1		X	1	3
18		X	1			1	2
19	1		2	X		1	4
20	2		X	1			[3]
21	1		1	X			2
22				X			0
23	X		1		1		2
24			2	X	1		3
25			2		X	1	3
26	X		[8]	1			9

64

Table 7   - Continued

No. Item	Choices						Total
	A	B	C	D	E	O	
27	X	2			1		3
28		X				1	1
29				X	3	2	5
30	2		X			1	3
31	X	[4]		1	1	3	9
32	2	X				1	[3]
33	X			[8]	1	1	10
34	X			[4]		1	5
35		3		X	3	1	7
36	X	[9]	1	1		1	12
37		X		[10]		2	12
38	1	4	[7]	X		2	14
39		3	3	X	1	2	9
40	X	[5]			2	5	12
41	1	3		[6]	X	3	13
42	[10]		X		1	3	14
43	[8]	1		X		3	12
44	1	[4]	X		1	7	13
45			[10]	X		6	16

KEY: as for Table 6

Table 8 59

ERROR ANALYSIS OF MULTIPLE-CHOICE
ANSWERS IN SUBTEST V

No. Item	Choices						
	A	B	C	D	E	O	Total
1	[4]	X	1	2		2	9
2		1			X	1	2
3	X		1	1	[3]	1	6
4		2	X	1		2	5
5	[3]		1	X	1	2	7
6	1	X	2	1		1	5
7	2	X					2
8	1		X			1	2
9			1	1	X	1	3
10	1	1	1	2	X		5
11	1	[3]		X	1	2	7
12		X	1	1	[3]	2	7
13	1	2	X			2	5
14		X		1	1	.	2
15	X		1		1		2
16	[3]	1	1	X		1	6
17		X		1	1	1	3
18	2	[3]			X		5
19	[3]			X	1		4
20	3	2	2	X	1	1	9
21		1	X	[3]			4
22	1		X	2			5
23	2		[4]		X	1	7
24	X		1	[3]	2		6

KEY: as for Table 6

Figure 1

COMPARATIVE RAW-SCORES FOR SUB-TESTS AND TOTAL
(as %ages), arranged according to overall rank ordering
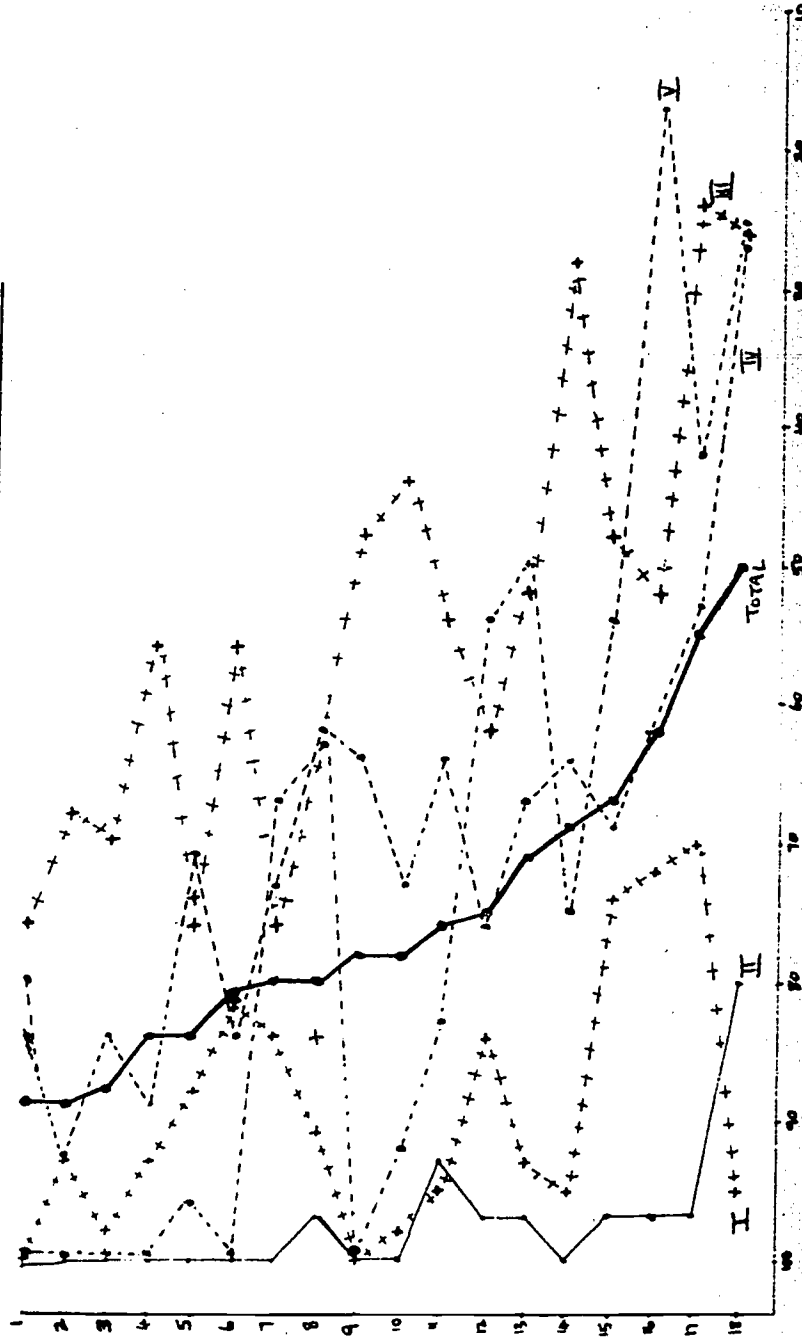
Figure II

COMPARATIVE RANK ORDER CURVES FOR RAW-SCORE TOTALS AND 5 SUB-TESTS. (Reduced to %ages.)
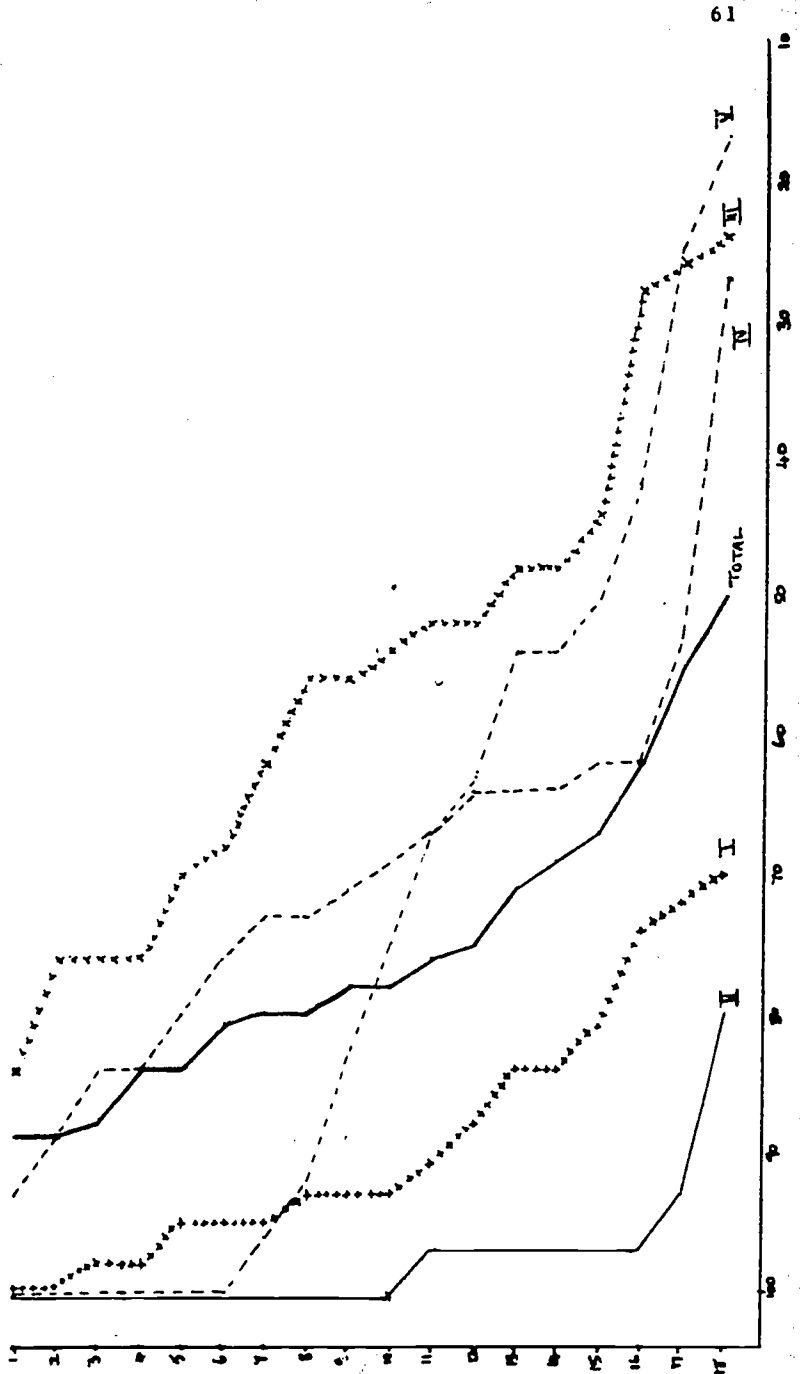
An Item Analysis of Parts III and IV

I. Shepherd

Since Parts III and IV of the MLAT appeared to give a more nearly normal curve than the other parts of the test, it was thought worthwhile to carry out an item analysis of these parts to discover what were the relevant factors operating. This analysis is related to Lee's paper above in that some of the operations he used on a small sample are applied to the total number of test papers, and that his sample results are compared with the results of the full analysis.

The 594 papers (this total is hereafter referred to as N) were divided into six equal groups on the basis of total test score. The answers of each group to each question in Parts III and IV were then analysed to give the following information:-

1. The number of students who chose the correct alternative.

2. The number who chose each of the incorrect alternatives.

3. The number who omitted to answer the question but answered at least one later question.

4. The number by whom the question was unattempted; i.e. that question and all subsequent questions were unanswered.

From this information the following factors were isolated for each question:-

a) Facility:- The percentage of correct

answers (hereafter referred to as C).

b) Difficulty:- The percentage of positive
errors (hereafter referred to as E), and the
percentage of omissions (hereafter referred to
as O). O may also be related to speed.

c) Speed:- The percentage of unattempted
questions (hereafter referred to as U).

d) Discrimination:- C of the two highest
groups minus C of the two lowest groups
expressed as a percentage. This figure was
analysed to show the contributions of diffi-
culty and speed to discrimination.

e) Distraction:- The percentage of E
represented by each of the incorrect alter-
native answers, especially the most frequent-
ly chosen one - the "major distractor".

These factors and their operation in each of
the two sub-tests are taken up below and illustrated
in Tables 1 - 4 and Figures 1 - 4. In addition
questions shown by the analysis to be of particular
interest are individually discussed.

I. Sub-test III

a) Facility:- C fluctuates between 96% and
35% in the first fifteen questions, in most
questions falling between 90% and 50%. Thereafter
C decreases steadily from 80% at question 16 to 4%
at question 38 and remains at this low level until
the end. (See Fig. 1.)

b) Difficulty:- E and O correlate with C
for the first fifteen questions and then remain at
a uniformly low level. (See Fig. 1.) Only fifteen
of the fifty questions have a figure for E of over
5%, and of these figures only three exceed 25%.
The equivalent figures for O are twelve and nil.
(See Table 1.)

c) Speed:- Sub-test III is highly speeded
(5:3) and so there is a figure for U as early as
question 10. It rises sharply from 2% at question
13 to 81% at question 33, and then more gradually
to 95% at question 42. Thereafter it remains
above 95%.

70

As E and O cease to correlate with C after
question 15, U begins to correlate and retains the
correlation until the end of the sub-test. (See
Fig. 1.)

d)   Discrimination:- This figure fluctuates
between 2% and 33% in the first thirteen questions
and then rises more or less steadily to 51% at
questions 23 and 24.   Thereafter it decreases
steadily to 8% at question 38 and then remains at
a low level. (See Fig. 1.)

It is clear from the observations made under
headings a), b) and c) above that sub-test III
falls into two parts: before speed becomes the
dominant factor, and after that point; the division
comes at or about question 15.

For the first fifteen questions there is some
positive correlation between discrimination and E
and O, indicating that discrimination is by diffi-
culty. This is apparent when E and O are taken
separately, and is even clearer when they are
combined, particularly if questions 14 and 15 are
discounted as already considerably affected by
speed. (See Table 2 and Fig. 2, Nos. 3,4 and 5.)

There appears to be no correlation between
discrimination and distraction. (See Fig. 2, No.2.)

Of the thirteen questions discriminated mainly
be difficulty, only question 12 has a discrimination
figure of over 30%, (a lower limit than many test
constructors would be willing to accept).

When the speed factor becomes dominant after
question 13 there emerges a very clear correlation
between discrimination and U. It is a non-mono-
tonic correlation, however, so that questions with
a low figure for U (early in the sub-test) and those
with a high figure for U (late in the test) show
poor overall discrimination, while questions with
a figure for U between 20% and 75% (questions 19 -
31) have good discrimination figures - between 34%
and 51%. (See Fig. 2, No. 1.)

[It is perhaps worth noting that questions
which do not appear to follow the curve in Fig.2,
No. 1 are those where difficulty and speed both
contribute significantly to discrimination. (See
Table 2.)]

71

The nature of the discrimination/unattempted
curve shows an essential feature of a speeded test.
There must be a large number of questions to force
speed, and so the later questions only discriminate
among the best subjects, not overall. Conversely
the early questions will eliminate the weaker
subjects, but again will not discriminate over the
whole range. Moreover, in a speeded test difficulty
only operates decisively, as here, in the very
early questions, before speed becomes dominant.
Even then, it may be argued, it is speed which forces
errors, and, more obviously, omissions - there are
few omissions in a non-speeded test.

The discrimination pattern of Sub-test III and
its contribution to the discrimination pattern of
the MLAT as a whole may be seen more clearly when
the figures for C in the six groups are examined.
Generally they are in descending order correspond-
ing to the order of the groups on the total test
score. Negative discrimination is not significant;
only thirty-nine out of two hundred and fifty
group discrimination figures are negative, and
twenty-four of these are in the last fifteen
questions where the figures for C are too small to
be relevant. In any case the average negative
discrimination in all these cases is less than 2%.

In the early questions the figures for C in
the six groups descend by fairly even steps, with
a slightly smaller difference between the two
middle groups. As the effect of speed makes itself
felt, the difference between groups 5 and 6 becomes
prominent for a short time, and then the difference
between groups 1 and 2, with a smaller increase in
the difference between the two middle groups. As
the figure for C decreases the difference between
groups 1 and 2 becomes increasingly dominant until
by the end of the Sub-test it represents almost the
whole of what little discrimination there is.

e) Distraction:- As stated above (p.63),
only fifteen of the fifty questions in Sub-test III
have a figure for E of over 5%, and it seems
reasonable to confine consideration of distraction
to these questions.

As Lee noted (p.40), the multiple-choice
alternatives do not distract equally. Of the sixty
distractors in the questions considered, two are

90 - 100% of E, two are 80 - 89% of E, four are
70 - 79% of E, two are 60 - 69% of E, two are 50 -
59% of E, three are 40 - 49% of E, two are 30 - 39%
of E, five are 20 - 29% of E, twelve are 10 - 19%
of E, and twenty-six are under 10% of E. (See
Table 1.)

Lee's comment that the speed factor resulted
in a preponderance of errors occurring before the
correct item (p40) is borne out by the full analysis,
but the effect is less marked. 68% of errors occur
before the correct item in the full analysis, com-
pared with 90% in the sample. The significance of
the effect is in any case reduced by the fact that
there are 112 choices before the correct item com-
pared with 88 after it.

f)   Individual Questions:- Lee (pp. 40,41)
picks out for comment seven questions having a
strong major distractor or a high number of
omissions. It would seem appropriate to begin with
these seven, commenting on Lee's observations in
the light of the full analysis, and then to
examine other questions, which did not appear
significant in the sample analysis but do appear so
in the full analysis.

1.   Question 2 : rgument.

The two analyses agree on the major distractor,
regiment; in the full analysis it accounts for 92%
of E where E is 29% of N. This is the highest
major distractor and the third highest figure for
E in Sub-test III.

2.   Question 4 : nme.

The two analyses agree on the major distractor,
sea-animal; in the full analysis it accounts for 67%
of E where E is 39% of N. In the full analysis,
however, there is another distractor, rush ess,
accounting for 29% of E. This question has the
highest figure for E in Sub-test III.

3.   Question 8 : ple

In the sample analysis this question has the
highest figure for E, but in the full analysis it
has only the fifth highest figure for E, and also
has a low figure for E, and for discrimination.
In the full analysis the major distractor is tree
(from apple?).

4.   Question 12 : ndkat.

    The two analyses agree on the major distractor,
species of feline; but in the total analysis it
accounts for only 41% of E where E is 11% of N.
This is clearly a puzzling question, however, as
the figure for O is 25%, the highest in Sub-test
III.   [The figure for the omission component of
discrimination is also the highest in Sub-test III.
(See Table 2.)]   The difficulty lies, perhaps, in
the number of steps that have to be made to reach
the required word. The letter-name 'n' has to be
used instead of its sound, but it has to be changed
to /in/; a syllable and a vowel, /i/ or /ə/, have
to be inserted between 'd' and 'k' with no clue
given; and finally 'a' has to be given its letter-
name sound /ei/, which it would never have in such
a position in English spelling.

5.   Question 14 : snser.

    The two analyses agree on the major distractor,
respond; in the full analysis it accounts for 77%
of E, where E is 36% of N. This is the second
highest figure for E in Sub-test III, and this
question also has the second highest figure for O
(24% of N). It is, therefore, the most difficult
question in Sub-test III, but not a good one, as
the discrimination is only 18%, of which difficulty
only accounts for ?%, so that it is clear that both
good and weak subjects were confused by the question.
Like question 12 it requires vowel changes, letter-
name sounds, and the breaking of English spelling
rules to arrive at the correct word. The choice of
respond is perhaps explained by the reading of
snser as anser (answer) as the pressure of speed is
beginning to be felt.

6.   Question 22 : kurs.

    The two analyses agree on the major distractor,
race-track, but as speed is an important factor by
this point, the figures for E and O in the full
analysis are very low (9% and 2% of N respectively).
As Lee says, the error represents a confusion of
course and curse. The confusion is perhaps
increased by the fact that in the only other
question in which 'u' occurs (question 2, rgumnt)
it has the sound / ju/, and the general inconsist-
ency of vowel sound/symbol values has precluded
the reading closest to orthography.

⁻.   Question 50 : pes.

The two analyses agree on the major distractor, vegetable; although in both, of course, very few subjects reached the last question.

8.   Question 7 : kwam.

This has the fourth highest figure for E and the third highest for O in Sub-test III, 18% and 19% respectively. There are two almost equal major distractors (41% and 38% of E), citrus fruit (yam?) and dwelling (wigwam?) but the comparatively even distribution of errors indicates a general difficulty rather than the effect of a strong distractor. This difficulty possibly stems from a combination of the comparative rarity of the lexical item qualm, and the problem of assigning a value to 'a'. This is the first time 'a' occurs in the Sub-test, but from the previous occurrence of other vowels, subjects would tend to hesitate between the letter-name value, /ei/, and the English spelling rule value /æ/; the value /a/ would not readily suggest itself. (In fact, apart from this question, 'a' always has the value /ei/ or /æ/, except when followed by 'r'.)

9.   Question 10 : thnkfl.

This has low figures for E and O but an extremely high major distractor, thoughtful (91% of E). The subjects choosing this alternative presumably take the letter-name of 'n' and change it to /in/ (as required in question 14) instead of to /æn/, thus manufacturing a (supposedly American-English?) word thinkful.

10.   Question 11: knfrns.

This has fairly high figures for E and O (16% and 11% respectively) and a high major distractor, kind of tree (conifers?) (86%). It also has the second highest figure for discrimination by omission (23%). (See Table 2.) Here there is no clue to the vowel needed between 'k' and 'n', but those choosing the major distractor were not led astray by this; they chose the right vowel but ignored the second 'n' to arrive at conifers.

11.   Question 28 : kataklizm.

This has a high E figure for so late in the
Sub-test (10%) and a high major distractor (88% of
E), chemical reagent (catalyst?).   kataklizm seems
close to a "simplified spelling" version of
cataclysm, so it was perhaps the unfamiliarity of
this lexical item which led subjects to choose an
incorrect alternative occurring after the correct
one in the list of multiple-choice answers.

II.   Sub-test IV

a)   Facility:-   C decreases from 96% at
question 1 to 10% at question 45, with individual
fluctuations. (See Fig. 3.)

b)   Difficulty:-   O is generally not signifi-
cant in the Sub-test, as would be expected in a
test which is not highly speeded; the highest
figure for O is 6% of N.   As in the sample analysis
(Lee p. 41), E is generally high.   All except two
questions have a figure for E of over 10%, and of
these thirty-two are over 25%; eleven exceed 50%.

The general decrease and the fluctuations of
C correlate negatively with E except for the last
few questions, where speed plays some part in the
answer pattern.   (See Fig. 3.)

The increase of E was noted in the sample
analysis (Lee p. 42) but only as occurring at the
end of the Sub-test.   No evidence appears in the
sample analysis of an actual increase in item
difficulty, but as in the full analysis the increase
of E is apparent from the beginning of the Sub-test,
it would seem that there is in fact a grading of
item difficulty.

c)   Speed:-   Sub-test IV is not highly speeded,
but the appearance of a figure for U at question 22,
rising to 37% at question 45 shows that speed is an
increasingly important factor in the second half of
the test.   The figure for U exceeds C for the last
four questions, although it never exceeds E.

The effect of this is apparent in the distortion

76

of the correlation pattern of E and C after question
37. (See Fig. 3.)

    d)   Discrimination:- This figure increases
gradually from 8% at question 1 to 45% at question
34, with fluctuations. Thereafter it decreases to
10%, again with fluctuations. (See Fig. 3.)

    Of the 45 questions in the Sub-test, only
eleven have a discrimination figure of over 30%.
This is a small number for a Sub-test which is not
highly speeded, and where, therefore, all questions
would be expected to contribute to discrimination.

    There seems to be some positive correlation
between discrimination and difficulty for the first
thirty-four questions. (See Fig. 4, No. 3.) With
the appearance of a figure for U at question 22,
there is also a strong positive correlation between
the U component of discrimination and U; both are
rising - 'oidly. (See Fig. 4, No. 5.)

    After question 34 the positive correlation
between the U component of discrimination and U
continues, but is negatively accelerated. (See
Fig. 4, No. 5.) At this point also,.the E com-
ponent of discrimination falls sharply (with no
clear correlation with E) and for the last few
questions is a minus quantity. This counteracts
the effect of the U component, leaving the
correlation of the total discrimination figure with
U a negative one for these questions, (see Fig. 4,
No. 4) and establishing a non-monotonic correlation
pattern for the Sub-test as a whole. (See Fig. 4,
No. 1.)

    There is no correlation between discrimination
and distraction. (See Fig. 4, No. 2.)

    As with Sub-test III, an examination of the
figures for C in the six groups shows the contribut-
ion of Sub-test IV to the overall discrimination
pattern of the MLAT. As in Sub-test III these
figures are generally in descending order, although
in this case negative discrimination is slightly
more significant. Thirty-three out of two hundred
and twenty-five group discrimination figures are
negative, the average of these figures being 4%.

77

In contrast with Sub-test 111, the relative
differences between the figures for C in the groups
do not show any pattern of change through the Sub-
test, although there are considerable individual
fluctuations.

e)   Distraction:- As all questions except
one have a figure of E of over 5%, the distraction
pattern is clearer than in Sub-test III.

Again, the alternatives do not distract
equally; of the 176 distractors in the forty-four
eligible questions, one is 90-100% of E, seven are
80-89%, eight are 70-79%, eight are 60-69%, twelve
are 50-59%, eight are 40-49%, twelve are 30-39%,
eighteen are 20-29%, thirty-six are 10-19%, and
sixty-six are under 10% of E.   (See Table 3.)

In this Sub-test there is a slight preponder-
ance of errors occurring before the correct answer
(53%) where the number of choices is almost
equally divided (91 before the correct answer and
89 after it).

f)   Individual Questions:- As in the
corresponding section of the examination of Sub-
test III, the questions selected for comment in the
sample analysis will be discussed first, followed by
questions suggested by the full analysis.

1.   Question 9.

It is seen as significant in the sample
analysis that two out of eighteen subjects omitted
a question comparatively early in the test.   In the
full analysis, only eight subjects omitted it.   The
other questions with -ing stimuli do not seem
particularly significant in the full analysis,
except for question 43.   (See below.)

2.   Question 11 : They named him BILL

Because of his military success during
A (2%)

the Civil War, the people made Grant
B(1%)   C (1%)   D (96% of E)

president of the United States.
E (X)

E is 29% of N.   The two analyses agree on the

major distractor.

As Lee says (p.43) the general similarity
between two proper nouns distracts attention from
the difference of function between them.

3.   Question 12 : The company owns every
      substantial PIECE of property in the town.

      Before the dawn of history, men were
                 A(71%)   B(1%)

raising corn very much like what we grow today.
       C(X)   D(25%)   E (2%)

E is 37% of N.   The two analyses agree on the
major distractor. Again, as Lee says, surface
similarity overrides functional difference.

4.   Question 16 : My finger became SWOLLEN
      from the infection.

      The child grew strong from the healing
                      A (X)   B (11%)

sunshine.

      The high wall was nearly hidden from view
           C(2%)   D(85%)

by the foliage.
         E (2%)

E is 22% of N.   The two analyses agree on the
major distractor. Perhaps there is more justifica-
tion for the error as there is frequently ambiguity
of function with words with -en (or -ed) endings.

5.   Question 26 : Do AS I say

      Although the weather report predicted
      A(X)

clear skies for today, it rained all day.
B (6%)   C(60%)   D(13%)   E(21%)

E is 49% of N. This question has the highest
figure for O, 6%. The two analyses agree on a high
figure for E. Two of the other conjunction items
mentioned in the sample analysis do not appear
particularly significant in the full analysis, but
question 34 does. (See below.)

6.   Questions 31 and 40 :

31.   Which one do <u>YOU</u> think it is?

<u>That one</u> may belong to <u>me</u>.
   A(X)   B(31%)

<u>Please</u> pay <u>me</u> before going on <u>your</u> trip.
  C (4%)   D(54%)   E(11%)

40.   Which colour do <u>YOU</u> like best?

This <u>one</u> suits <u>me</u> better than the <u>other</u>.
     A(X)   B(63%)   C(4%)

It makes no <u>difference</u> to me.
              D (5%)   E(28%)

   E is 57% of N in question 31, and 50% of N in
question 40.

   The two analyses agree on the distracting effect
of similarity of parts of speech. In the full
analysis the effect of the major distractor and the
significant minor distractors depends on pronoun
being matched with pronoun, regardless of function.

7.   <u>Question 36</u> : A <u>NUMBER</u> of people applied
      for the position.

   <u>I</u> find <u>many</u> <u>candidates</u> <u>who</u> cannot offer
   A(X)   B(77%)   C(16%)   D(3%)

more than <u>two</u> years'experience.
          E(5%)

   E is 67% of N, the highest figure in the Sub-
test. As Lee says, meaning plays a part, overriding
functional considerations, and making <u>many</u> the major
distractor in both analyses.

S.   <u>Questions 37 and 45</u> :

37.   His wife bought <u>HERSELF</u> a new hat.

Why won't <u>you</u> tell <u>me</u> <u>more</u> about
          A(2%)   B(X) C(6%)

<u>yourself</u> than <u>you</u> did yesterday.
  D(89%)   E(4%)

E is 52% of N.

45.   The child hurt <u>HIMSELF</u>.

Although I <u>myself</u> would do that by <u>myself</u>,
            A (6%)   B (6%)

Mary gained <u>herself</u> the <u>help</u> of some of her
            C (83%)   D (X)

<u>classmates</u>.
    E (5%)

E is 53% of N.   Again the surface similarity
overrides the functional difference, making <u>herself</u>
the major distractor in both analyses.

9.   <u>Questions 38 and 44</u> :

38.   <u>WHAT</u> is this?

I do not know <u>what</u> book you want.
              A(4%)

To <u>whom</u> do these belong?
   B(19%)

<u>Which</u> fellow is your brother?
C(75%)

<u>Those</u> are <u>mine</u>.
D(X)   E(3%)

E is 54% of N.

44.   There is no <u>POINT</u> in going ahead.

When the <u>light</u> changed, he stopped the
         A(17%)

<u>car</u>.
B(44%)
     A <u>river</u> flows <u>down</u> to the <u>sea</u>.
       C (X)   D(30%)   E(8%)

E is 46% of N.

The two analyses have the same major distract-
ors. In question 38, <u>which</u> is perhaps seen as the
most likely match, as <u>it</u>, like <u>what</u>, precedes <u>is</u>
in a question, and in fact could occur alone without

changing the meaning.

In question 44 there is apparently difficulty in equating the post-verbal noun in a there is sentence with a pre-verbal subject noun, river and so the post-verbal noun car or the post-verbal adverb down are chosen.

10. Question 41 : We plan to take IT today.

On the chance that he would see us,
A (2%)   B(44%)

we took steps to put up a beacon.
C(2%)   D(52%)   E(X)

E is 45% of N. The two analyses agree on the major distractor. As Lee says, the major distract-or matches it with steps by their relationship to take, ignoring the different status of take in the two sentences. The minor distractor links it with us, presumably because they are both pronouns, and both objects of verbs.

The large number of questions with a high figure for E means that a large number of questions could be commented on, but a representative sample of difficult questions with strong major distract-ors can be obtained by considering those questions with major distractors accounting for over 70% of E, where E is over 25% of N. This sample provides .llustrations of the various difficulties encountered by the subjects and the typical errors made. Of the thirteen questions satisfying these criteria, six were discussed in the sample analysis and have therefore already been commented on; the other seven are now considered.

11. Question 14 : SEVERAL were absent from the meeting.

In spite of the many proposals which
A(4%)   B(73%)   C(11%)

were made, only one could be adopted.
D(11%)E(X)

E is 31% of N. This question is comparable with question 36 in that similarity of meaning of quantifiers overrides difference of function.

12. Question 24 : He drove **FROM** Boston to
New York.

To be safe, he decided to buy spare
$\overline{A(10\%)}\overline{B(5\%)}$ $\overline{C(77\%)}$

parts for any emergency.
      $\overline{D(X)}$ $\overline{E(8\%)}$

E is 26% of N. Here a preposition, from
following a verb is matched with an infinitive
marker, to following a verb.

13. Question 25 : He nailed the board **TIGHT**
against the house.

He always did the job well.
   $\overline{A(4\%)}$ $\overline{B(2\%)}\overline{C(87\%)}$

He poured the pail full
            $\overline{D(7\%)}\overline{E(X)}$

E is 34% of N. The "adjective of result",
tight is matched with the adverb well. Both are
in the same position, following the object of a
transitive verb.

14. Question 28 : The weekly meeting, usually
held on Friday night, is a fixed **ACTIVITY**
of the Scout program.

Washington was the first president of the
  $\overline{A~(8\%)}$ $\overline{B~(X)}$

United States; he refused the crown that some
  $\overline{C~(4\%)}$ $\overline{D(74\%)}$

of his admirers wanted him to have.
               $\overline{E(13\%)}$

E is 35% of N. The noun head of the subject
complement following is is equated with the object
of refused.

15. Question 30 : **NONE** was more curious to
solve the riddle than I.

The government's first task was to check
    $\overline{A(7\%)}$ $\overline{B(6\%)}$ $\overline{C(X)}$

the prescriptions written by the doctors.
    $\overline{D(7\%)}$ $\overline{E(12\%)}$

E is 62% of N. None, subject of was is matched
with government's, one of the modifiers in a noun
phrase, subject of was. This is the second highest
figure for E in the Sub-test and the major distract-
or therefore represents a far greater percentage
of N than C does, which seems strange in view of the
apparently obvious difference between none and
government's. It is possible that government is
seen as the "situational" subject of the sentence
and therefore matched with none.

16. Question 34 : I will buy a car WHEN I get
     the money.

     After you left last night, most of the
     A(X)   B(2%)   C(15%)

     students remained until the end.
                    D(81%)   E(1%)

E is 34% of N. When, a conjunction linking
main clause with a subordinate clause following it
is matched with until, a preposition introducing
modifying phrase. The similarity of position and
meaning has overridden the functional difference.

17. Question 43 : Some people enjoy EATING
     clams on the half-shell.

     Hacking his way through the teeming
     A(82%)   B(3%)

     jungle, he found abundant evidence of the
                       C(4%)   D(X)

     vanished civilisation.
     E(5%)

E is 54% of N. Two -ing forms are equated.
Both have objects but their function is quite
different.

_____

Lee concluded (p.43)that the most general
cause of error was the tendency for subjects to
pay attention to general and surface similarities
rather than functional ones in the particular sent-
ences. The full analysis, illustrated in the
questions discussed above, confirms this. By far
the most common form of this error was the matching

84

of parts of speech regardless of function, but there
were other forms, such as matching by phonological,
orthological or morphological similarities, and
matching by position in relation to other items.
Differences between the stimulus sentence and the
response sentences in order of sentence elements,
and the occurrence of two-clause sentences obviously
caused difficulty.

Conclusion

Following Lee's pilot study, the full analysis
has indicated the factors at work in a speeded and
a non-speeded Sub-test of the MLAT and the relation-
ships between them. It has also shown the pattern
of errors in each Sub-test as a whole, and specific
difficulties in particular questions. It is hoped
that the information gained from this study, part-
icularly that related to discrimination, may be of
help to future test-constructors in designing
suitable aptitude tests for use at university
entrance level in Britain.

85

Tables & Figures

referring to the above article

Table 1 - ERROR ANALYSIS OF MULTIPLE CHOICE ANSWERS IN SUB-TEST III.

Question Number	a	%L	b	%L	c	%L	d	%E	e	%L	E	%N	O	%N	U	%N	C	%N
1	43	46	25	27	18	19	X	-	7	8	93	16	57	10	-	0	444	75
2	158	92	7		6	3	1	1	X	-	172	29	34	6	-	0	388	66
3	45	52	X		9	10	24	28	8	9	86	14	19	3	-	0	489	83
4	155	67	4	2	X	-	5	2	67	29	231	39	35	6	-	0	328	56
5	X		10		3		X		1	0	19	3	72	12	-	0	503	85
6	41	-	19	10	5	14	11	31	0	0	35	6	61	1	-	0	553	93
7	8	38	14	54	13	41	X	-	0	8	107	18	113	19	-	0	374	63
8	5	16	5	10	33	66	X	-	4	8	50	9	73	12	-	0	470	79
9	52	91	6		X		6	4	9		26	4	19	3	1	1	548	92
10	6	6	1		2	4	2	4	X	-	57	10	8	1	4	1	525	89
11	17	27	4	4	81	86	X	-	3	3	94	16	67	11	7	1	426	72
12	3		X	-	26	41	10	16	10	16	63	11	149	25	11	2	371	63
13	104	77	X		X		3		2		9	2	6	1	12	2	567	96
14	47	77	2		16	7	11	5	23	11	214	36	143	24	32	5	205	35
15	30	71	X		X	-	2	3	10	16	61	10	80	14	53	9	400	68
16	5		X		10	24	3	5	0		42	7	11	2	66	11	475	80
17	5		X		5		X		1		14	2	33	6	95	16	452	76
18	5		14		3		5		3		25	4	46	8	115	19	408	69
19	X		7		4		X		0		16	3	7	1	3	22	438	74
20	8		5		2		X		2	2	17	3	8	1	152	26	417	70
21	1		X		21		4		2		28	5	8	1	172	29	386	65
22	1	2	6	12	5		39	76	X		51	9	10	2	197	33	336	57
23	0		X		5		6		5		16	3	10	2	223	38	345	58
24	2		3		3		X		1		9	2	12	2	250	42	323	54
25	1		2		X		10		1		14	2	11	2	269	45	300	51
26	3		2		12		X		1		18	3	6	1	301	51	269	45

Table 1 - Continued

Question Number	a	%E	b	%E	c	%E	d	%E	e	%E	E	%N	O	%N	U	%N	C	%N
27	3	0	5	-	3	5	4	88	4	7	13	2	18	3	327	55	236	40
28	0	0	X	5	1	5	52		1		59	10	17	3	357	60	161	27
29	X		10		1		0		2		12	2	21	4	398	67	163	27
30	2		4		X		1		0		9	2	26		41	70	146	25
31	23		X		5		0		0		28	5	12		159	75	109	18
32	1		1		X		X		0		4	1	14		4	7	118	20
33	0		0		X		4		1		4		10	2	4		101	17
34	4		1		4		2		X		8	1	12	2	4		82	14
35	0		4		5		X		1		11	2	20	3	504	85	59	10
36	1		2		0		3		1		9	2	20	3	522	88	43	7
37	10		X		3		X		1		13	2	14	2	529	89	38	6
38	3		3		X		2		0		9	2	19	3	541	91	35	4
39	0		X		1		0		X		3	1	6	1	548	92	37	6
40	0		1		4		1		2		3	1	9	2	552	92	30	5
41	1		X		0		2		0		9	2	7	1	557	94	21	4
42	1		1		X		X		0		2	0	11	2	561	95	20	3
43	0		0		0		0		X		0	0	9		563	95	22	4
44	0		0		X		1		1		1	0	5	1	566	95	22	3
45	0		1		2		X		0		3	1	3	1	570	96	18	3
46	1		X		0		0		0		2	0	3	1	572	96	17	3
47	0		1		1		0		X		2	0	1	0	572	96	19	3
48	0		0		2		0		5		1	0	3	1	573	97	17	3
49	0		X		1		X		0		2	2	2	0	573	97	17	3
50	11		X		2		0		0		13	2	0	0	573	97	8	1

KEY: abcde = multiple choice answers. E = positive errors. Q = questions omitted.
U = questions unattempted. C = questions answered correctly. N = number of papers (594).

NOTE: 1. Percentages may not total 100 due to rounding off.
2. Percentages of E have been calculated only when E exceeds 5% of N.

Table 6 - ANALYSIS OF DISCRIMINATION IN SUB-TEST III.

Question Sub-test	E			O			U			C			
	H1/3	DF	L1/3	H1/3	DF	L1/3	H1/3	DF	L1/3	H1/3	DF	L1/3	DC
1	10	32	42	10	22	32	-	-	-	178	54	124	27
2	34	45	79	9	5	14	-	-	-	155	50	105	25
3	11	36	47	2	3	5	-	-	-	185	39	146	20
4	54	48	102	7	11	18	-	-	-	137	59	78	30
5	2	10	12	15	21	36	-	-	-	181	31	150	16
6	4	18	22	1	4	5	-	-	-	193	22	171	7
7	20	26	46	34	8	42	-	1	1	144	34	110	7
8	12	6	18	29	-3	26	-	1	1	157	4	153	15
9		16			10	11	-	1	1	196	26	170	15
10		13	24	1	5	6	4	4	4	186	22	164	11
11	17	29	46	10	23	33	5	5	5	171	57	114	29
12	7		31	29	34	63	8	8	8	162	66	96	3
13		7	7	1	2	3	8	8	8	197	17	180	9
14	63		77	50	-1	49	+	24	24	83	35	48	18
15	6	24	30	25	7	32	1	39	40	166	70	96	35
16	12	-1	11	1	7	8	3	42	45	182	48	134	24
17	-1		3	13	1	14	7	54	61	174	54	120	27
18	2	11	13	9	10	19	9	60	69	178	81	97	41
19	4	2	6	2	2	4	12	63	75	180	67	113	34
20	2	8	10	0	6	6	15	67	82	181	81	100	41
21	8	5	13	0	7	7	16	74	90	174	96	88	43
22	16	4	20	2	4	6	22	79	101	158	87	71	44
23	2	8	10	3	4	7	27	90	117	166	102	64	51
24	0	6	6	2	5	7	35	90	125	161	101	60	51
25	3	2	5	2	5	7	43	89	132	150	96	54	48
26	6	1	7	0	5	5	54	85	139	138	91	47	46

Table 11 - Continued

Question Number	E H1/3	E DF	E L1/3	O H1/3	O DF	O L1/3	U H1/3	U DF	U L1/3	C H1/3	C DF	C L1/3	DC
27	2	5	7	7	0	7	63	85	148	126	90	36	45
28	22	-8	14	5	1	6	75	79	154	96	72	24	36
29	1	-4	5	8	+3	5	91	75	166	98	76	22	38
30	2	2	4	13	-3	10	100	68	168	83	67	16	34
31	6	3	9	5	-1	4	115	60	175	72	62	10	31
32	0	3	3	6	-2	4	119	57	176	73	58	15	29
33	2	-2	0	2	3	5	131	48	179	63	49	14	25
34	3	-2	1	4	0	4	135	46	181	56	44	12	22
35	3	0	3	8	0	8	143	38	181	44	38	6	19
36	0	6	6	8	-2	6	154	28	182	36	32	4	16
37	5	-2	3	4	3	7	159	24	18.	30	25	5	13
38	3	-1	4	9	-4	5	167	19	186	19	16	3	8
39	0	3	3	0	5	5	172	15	187	26	23	3	12
40	0	2	2	3	2	5	173	15	188	22	19	3	10
41	1	-4	5	1	3	4	176	13	189	20	20	0	10
42	0	-1	1	1	5	6	179	11	190	18	17	1	9
43	0	0	0	2	3	5	181	9	190	15	12	3	6
44	0	-1	1	1	2	3	182	10	192	15	13	2	7
45	2	-2	0	0	2	2	184	8	192	12	8	4	4
46	1	-1	0	0	3	3	185	8	193	12	10	2	5
47	0	1	1	0	1	1	185	8	193	13	10	3	5
48	0	0	0	0	2	2	185	9	194	13	11	2	6
49	0	0	0	1	0	0	185	9	194	12	9	3	5
50	7	-4	3	0	0	0	185	9	194	6	5	1	3

KEY: H= high   L= low   DF= difference between high one-third and low one-third of N.
DC= discrimination; i.e, DF(of C) expressed as a percentage.

90

Table III - ERROR ANALYSIS OF MULTIPLE CHOICE ANSWERS IN SUB-TEST IV.

Question Number	a	%E	b	%L	c	%E	d	%E	c	%E	E	%N	O	%N	U	%N	C	%N
1	6	-	9	10	X	7	16	17	X	-	24	4	2	0	-	0	568	96
2	66	-	19	54	0	-	8	23	0	0	92	15	2	0	-	0	500	84
3	23	-	41	50	31	38	X	-	3	4	35	6	2	0	-	0	557	94
4	9	-	9	11	19	24	X	-	45	56	82	14	3	1	-	0	509	86
5	4	-	15	16	X	-	63	69	9	10	80	13	2	0	-	0	512	86
6	-	-	98	65	6	4	14	9	33	22	91	15	3	0	-	0	501	85
7	34	49	25	36	X	-	84	6	7	10	151	25	1	1	-	0	440	74
8	1	1	44	26	X	-	54	51	37	22	70	12	8	0	-	0	523	88
9	63	84	X	-	8	11	167	4	1	1	169	28	0	1	-	0	417	71
10	4	2	2	1	1	1	167	96	X	-	75	13	0	0	-	0	519	88
11	159	71	3	1	X	-	54	25	5	2	174	29	5	0	-	0	420	71
12	18	17	31	29	X	-	44	42	13	12	223	37	4	1	-	1	366	62
13	8	4	133	73	21	11	21	11	X	-	106	18	3	1	-	1	484	82
14	90	36	119	48	27	11	14	6	X	2	183	31	3	1	-	1	408	69
15	X	-	15	11	2	47	112	85	X	-	250	42	5	1	-	1	339	58
16	6	5	57	45	60	70	4	3	X	2	132	22	3	1	-	1	459	77
17	18	15	X	-	82	45	2	3	15	13	127	21	2	0	-	1	465	79
18	47	22	27	13	95	-	X	-	43	26	117	20	4	0	-	1	473	80
19	100	53	7	4	X	21	74	39	7	4	212	35	18	3	1	1	364	62
20	124	67	21	11	36	28	X	-	1	1	186	31	2	0	3	1	403	68
21	79	54	19	13	40	22	X	-	1	1	184	31	1	0	5	1	408	69
22	X	-	9	5	42	77	34	18	7	5	145	24	1	0	7	1	445	75
23	15	10	7	5	119	87	X	-	108	56	193	32	1	0	9	1	595	67
24	8	4	7	2	175	87	14	7	13	8	154	26	26	4	-	1	407	69
25									X	-	201	34	4	1	-	2	380	64

Table III - Continued

Question Number	a	% t	b	% E	c	% L	d	% E	e	% E	E	% N	O	% N	U	% N	C	% N
26	X	-	18	6	176	60	37	13	61	-	292	49	33	6	12	2	257	44
27	X	-	170	63	10	4	57	21	31	-	268	45	5	1	13	2	308	52
28	17	8	X	-	9	4	155	74	28	-	209	35	9	2	18	3	358	61
29	78	35	34	15	28	13	X	-	84	-	224	37	13	2	20	3	337	57
30	278	75	23	6	X	-	24	7	44	-	369	62	6	1	25	4	194	33
31	X	-	106	31	15	4	184	54	36	-	341	57	19	3	29	5	205	35
32	131	58	X	-	5	2	72	32	18	-	226	38	8	1	34	6	326	55
33	X	-	94	30	40	13	154	50	22	-	310	52	7	1	43	7	234	40
34	X	-	5	2	30	15	165	81	3	-	203	34	2	0	49	8	340	58
35	85	35	83	34	30	12	X	:	47	-	245	41	9	2	57	10	283	48
36	X	-	307	77	62	16	12	3	19	-	400	67	2	0	69	12	123	21
37	5	2	X	-	18	6	273	89	11	-	307	52	5	1	81	14	201	34
38	12	4	60	19	238	75	X	-	9	-	319	54	2	0	98	17	175	29
39	17	5	185	57	38	12	X	5	85	-	325	55	14	2	117	20	138	23
40	X	-	186	63	13	4	16	5	82	-	297	50	14	2	131	22	152	26
41	6	2	117	44	6	2	138	52	X	-	267	45	7	1	145	25	175	30
42	212	61	20	6	X	-	47	14	69	-	348	59	4	1	159	27	83	14
43	261	82	26	8	14	4	X	-	17	-	318	54	8	1	184	31	84	14
44	47	17	121	44	X	-	83	30	23	-	274	46	19	3	213	36	88	15
45	19	6	18	6	258	83	X	-	16	-	311	53	5	1	219	37	59	10

KEY: As Table I

Table IV - ANALYSIS OF DISCRIMINATION IN SUB-TEST IV.

Question Number	E			O			U			C			
	HI/3	DF	LI/3	HI/3	DF	LI/3	HI/3	DF	LI/3	HI/3	DF	LI/3	DC
1	11	50	61	0	1	1	-	-	-	197	15	182	8
2	4	21	25	1	0	1	-	-	-	186	50	136	25
3	9	37	46	0	2	2	-	-	-	194	23	171	12
4	6	48	54	0	3	3	-	-	-	189	40	149	20
5	11	45	56	1	-1	0	-	-	-	191	47	144	24
6	20	68	88	0	2	2	-	-	-	187	47	140	24
7	7	43	50	0	2	2	-	-	-	178	70	108	35
8	35	46	81	0	1	1	-	-	-	191	44	147	22
9	16	21	37	3	0	3	-	-	-	160	46	114	23
10	38	44	82	0	0	0	-	-	-	182	21	161	11
11	47	56	103	0	4	4	-	-	-	160	44	116	22
12	26	19	45	2	-1	1	-	-	-	151	60	91	30
13	31	53	84	0	2	2	-	-	-	170	18	152	9
14	74	10	84	2	0	2	-	-	-	167	55	112	28
15	40	4	44	0	2	2	-	-	-	122	10	112	5
16	18	59	77	0	2	2	-	-	-	158	7	151	4
17	30	27	57	0	3	3	-	-	-	180	61	119	31
18	48	54	102	5	3	8	-	-	-	168	30	138	15
19	41	34	75	1	0	1	1	1	1	145	57	88	29
20	36	58	94	0	0	0	1	1	1	156	35	121	18
21	25	46	71	1	-1	0	3	3	3	162	59	103	30
22	50	33	83	0	1	1	5	5	5	172	48	124	24
23	40	40	80	3	11	14	7	7	7	148	39	109	20
24	51	32	83	0	-4	-4	8	8	8	155	58	97	29
25										147	44	103	22

Table IV . Continued

Question Number	E H1/3	DF	L1/3	Q H1/3	DF	L1/3	U H1/3	DF	L1/3	C H1/3	DF	L1/3	DC
26	84	36	120	11	-3	8	-	10	10	103	43	60	22
27	54	69	123	2	-1	1	-	11	11	142	79	63	40
28	34	66	100	4	0	4	-	15	15	160	81	79	41
29	52	38	90	4	-2	6	b	17	17	142	57	85	29
30	91	51	142	3	-1	2	-	22	22	104	72	32	36
31	81	57	138	7	0	7	-	25	25	110	82	28	41
32	62	29	91	2	1	3	2	25	27	132	55	77	28
33	76	53	129	2	-1	1	4	29	33	116	81	35	41
34	43	56	99	0	1	1	4	33	37	151	90	61	45
35	72	8	80	9	-9	0	4	39	43	113	38	75	19
36	123	7	130	0	1	1	7	40	47	68	48	20	24
37	90	19	109	0	2	2	9	43	52	99	64	35	32
38	103	-3	100	0	1	1	14	43	57	81	41	40	21
39	96	18	114	4	2	6	18	45	63	80	65	15	33
40	83	31	114	3	0	3	21	47	68	91	78	13	39
41	103	-40	63	1	1	2	26	47	73	68	8	60	4
42	132	-33	99	1	0	1	29	47	76	36	14	22	7
43	108	-6	102	4	-2	2	34	51	85	52	43	9	22
44	100	-24	76	6	-2	4	44	54	98	48	28	20	14
45	119	-33	86	2	-1	1	46	54	100	31	20	11	10

KEY: as Table 11

94

Fig. 1   Sub-Test III

UNATTEMPTED

DISCRIMINATION

ERRORS
CORRECT
OMITTED
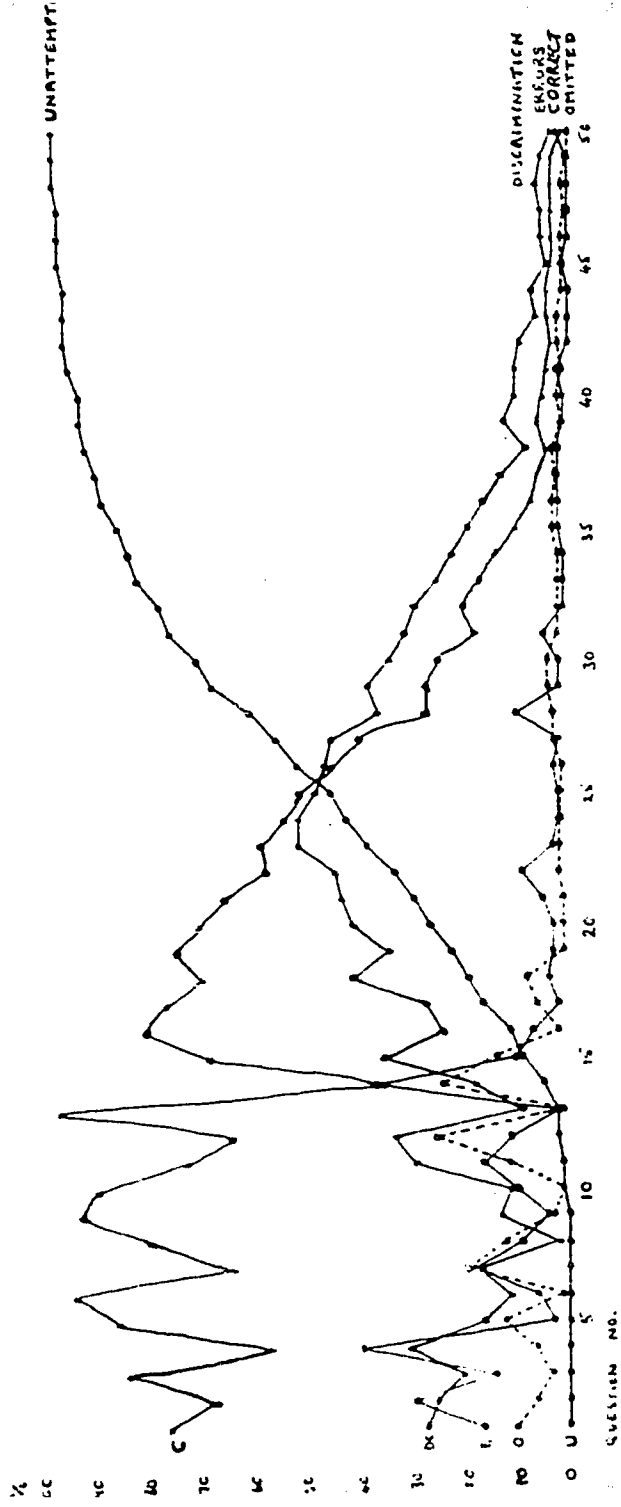
Question No.

Fig. 2   Sub-Test III   Correlation of Discrimination
with Other Factors
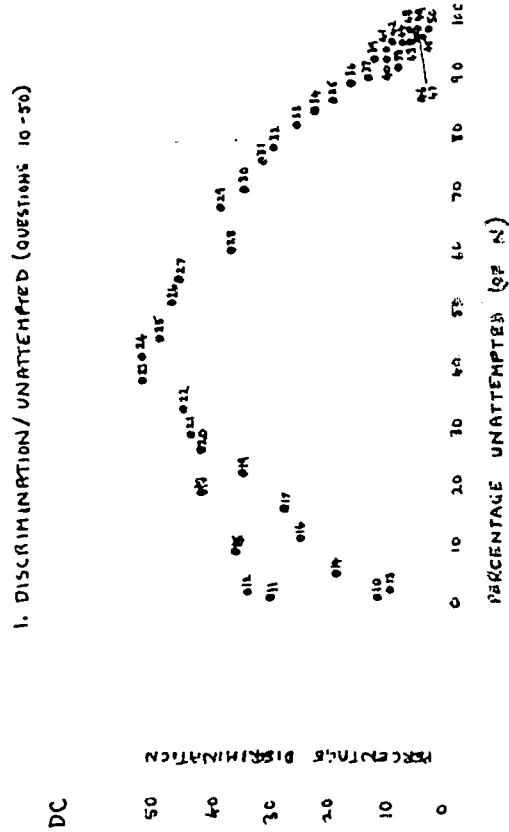
1. DISCRIMINATION/ UNATTEMPTED (QUESTIONS 10 -50)

PERCENTAGE UNATTEMPTED (OF N')

PERCENTAGE DISCRIMINATION

Fig. 2 Continued

2. DISCRIMINATION / MAJOR DISTRACTOR (QUESTIONS 1-15)

Ɔ⊂

50

40

30

20

:0

0

PERCENTAGE DISCRIMINATION

C   10   20   30   40   50   60   70   80   90   100

PERCENTAGE OF E REPRESENTED BY MAJOR DISTRACTOR

3. DISCRIMINATION / ERRORS (QUESTIONS 1-15)

DC

50

40

30

20

10

C

PERCENTAGE DISCRIMINATION

C   10   20   30   40

PERCENTAGE POSITIVE ERRORS (OF N)

Fig. 2 Continued

4. DISCRIMINATION/OMISSIONS (QUESTIONS 1-15)

5. DISCRIMINATION/ERRORS + OMISSIONS (QUESTIONS 1-15)

98
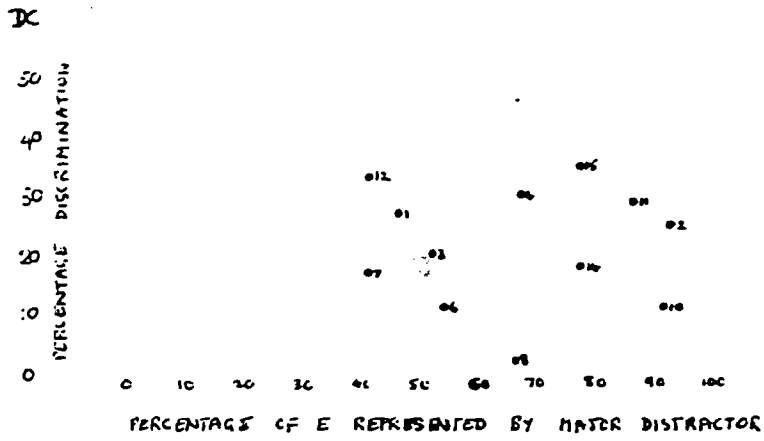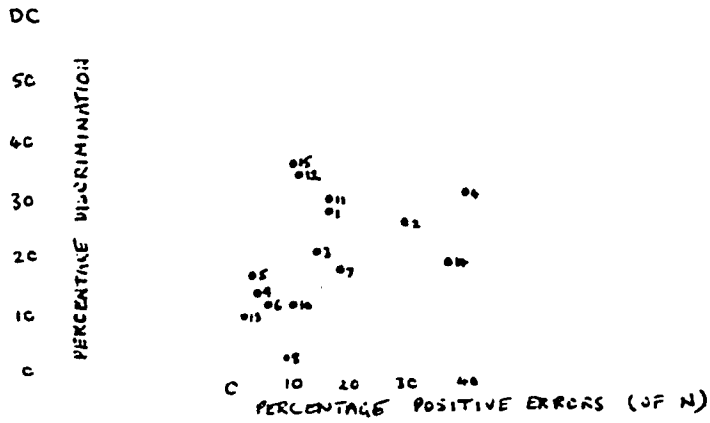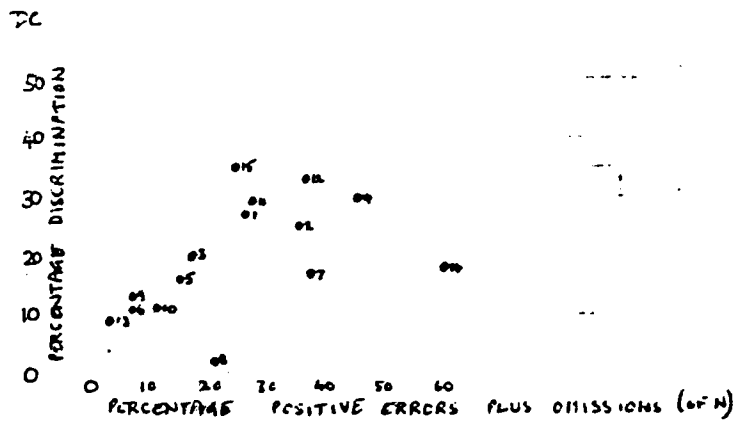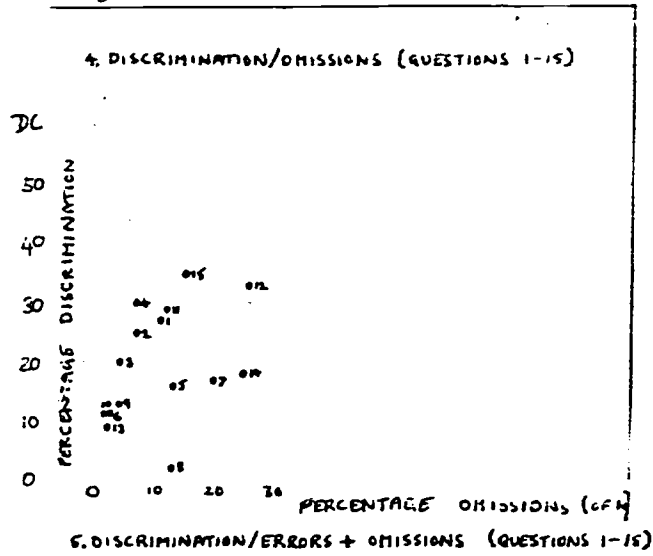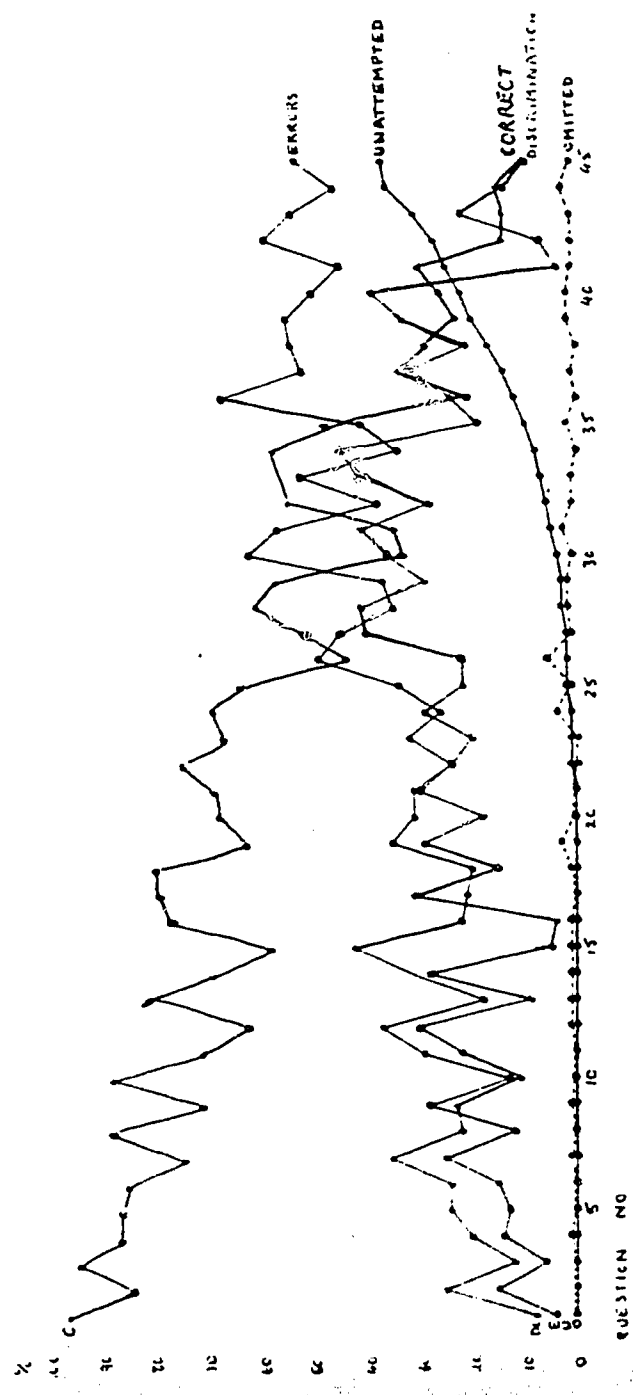
Fig.3   Sub-Test IV

Fig.1 Sub-Test IV Correlation of Discrimination with
Other Factors

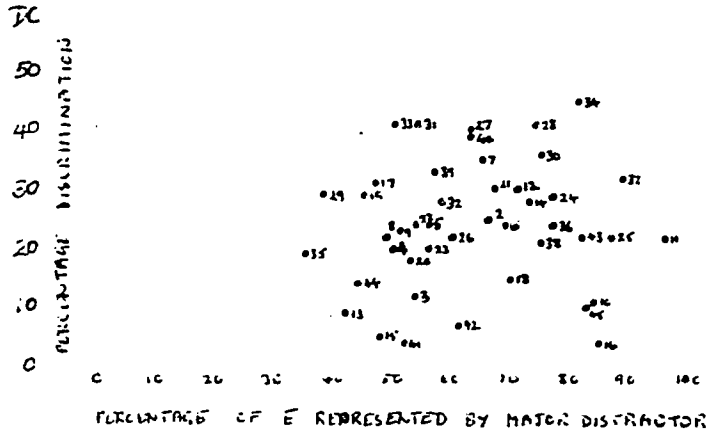: DISCRIMINATION / ERRORS + OMISSIONS + UNATTEMPTED (ALL QUESTIONS)
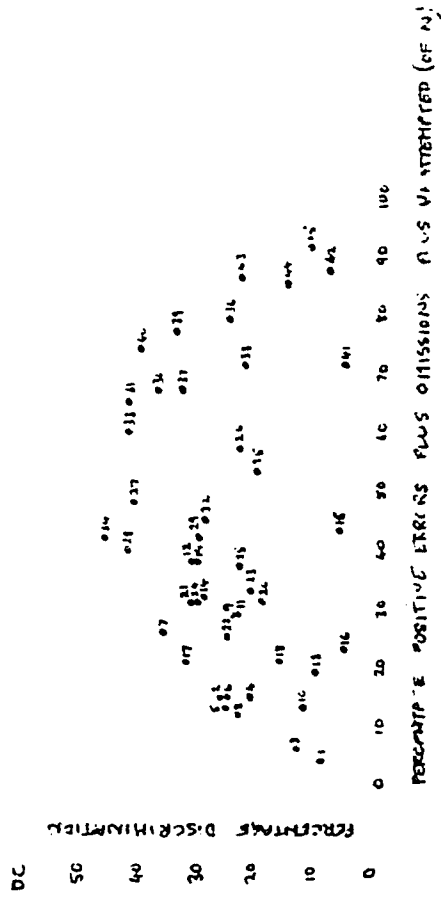
PERCENTAGE DISCRIMINATED

DC 50 40 30 20 10 0

0 10 20 30 40 50 60 70 80 90 100

PERCENTAGE POSITIVE ERRORS PLUS OMISSIONS PLUS UNATTEMPTED (OF N)

100

Fig.4 continued.

2. DISCRIMINATION / MAJOR DISTRACTOR (QUESTIONS 2-45)

PERCENTAGE OF E REPRESENTED BY MAJOR DISTRACTOR

3 DISCRIMINATION / ERRORS (QUESTIONS 1-36)

PERCENTAGE POSITIVE ERRORS (OF N)

101

Fig.4 continued

4. DISCRIMINATION/UNATTEMPTED (QUESTIONS 22-45)

5. DISCRIMINATION - UNATTEMPTED COMPONENT/UNATTEMPTED AND
X DISCRIMINATION - ERROR COMPONENT/ERRORS (QUESTIONS 22-45)
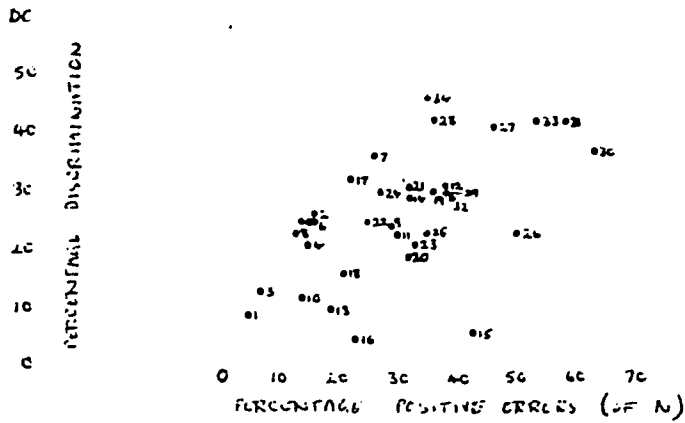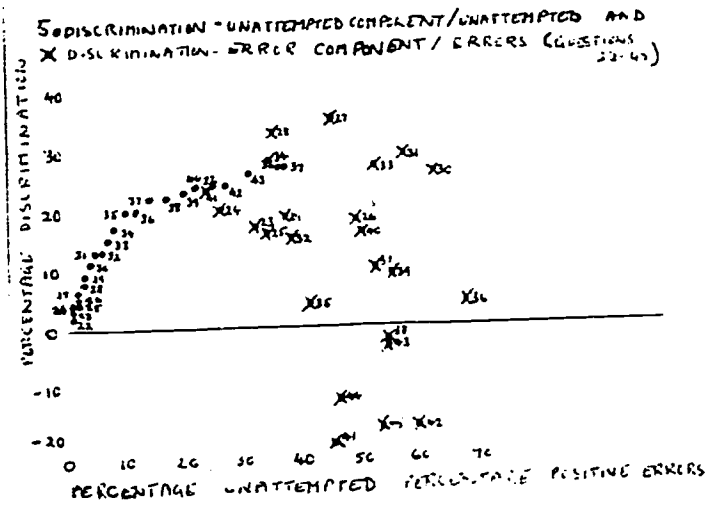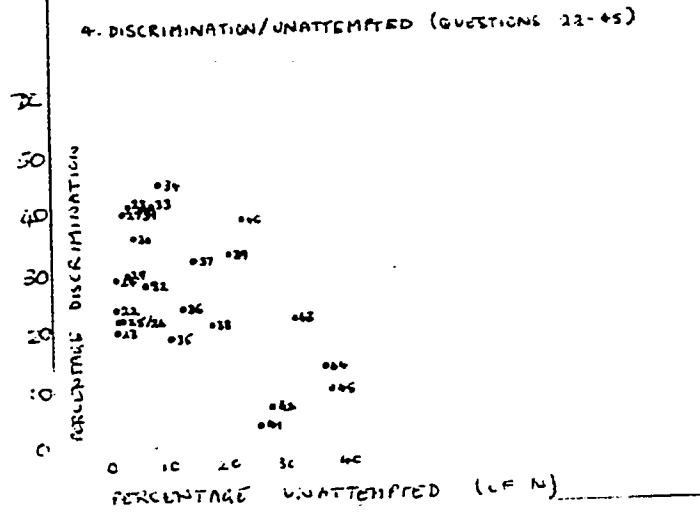
The Modern Language Aptitude Test Part V and the Learning of Foreign Vocabulary

M. H. T. Alford

The Scientific Language Project at the University of Essex developed a new method of learning to read foreign languages in which the memorization of vocabulary played a predominant part. Students learnt about 600 high-frequency words drawn from the target literature. The resulting high text-coverage of known words facilitated the learning of new words by deduction and ensured the maximum refreshment of those already memorized.

Grammatical instruction was carried out using vocabulary which had previously been learnt. This removed lexical distraction and made the examples much easier to remember. Since the examples were composed of high-frequency words, subsequent reading provided constant reminders of all the points which had been studied.

The MLAT Part V (Paired Associates) was used in designing the memorization procedures. It was necessary that people of all abilities should be able to master them and the test was used to pick suitable subjects to undergo the experiments.

Vocabularies of 600 words were memorized by about 50 individuals with scores ranging from 8 to 24. Five subjects who volunteered to continue memorization acquired over 1,000 words and all could have continued. Their MLAT scores ranged from 16 to 22. Three of them did Russian, one Spanish and one did both Hungarian and Rumanian.

Vocabulary memorization was carried out to a precise criterion. In learning Russian-English

paired associates (PAs), 28 were presented on
cards in each session. Time spent structuring the
background of each PA was kept to less than 50% and
students were advised to introduce attempted
remembering as early as possible. This was
performed with thirds of the pack shuffling at the
end of each imperfect test. Finally all the cards
were shuffled together and the process repeated
until the first correct recognition of all cards.

This procedure was designed to produce a
certain amount of underlearning on a 2 - 3 month
retention. The PAs which made up this underlearning
were ascertained by tests at interference intervals
of 140 and 270 PAs and the requisite relearning
carried out.

The initial session represented by far the
greater part of the memorizing effort and its smooth
functioning was essential to the success of the
method. The MLAT V was given to every student
before the course began. Instructions for the
procedures most suitable to each ability could then
be issued before problems were encountered.

The Test provided a very accurate prediction
of future performance. Memorizing the 28 Russian
PAs to the initial criterion mentioned above, the
following correlations were obtained:

Test Score (Max 24)	Average Memorizing Time (Mins)
22	25
18	35
16	40
12	55

Scores below 12 were advised to split memorizing
sessions into two or more parts. This enabled
them to achieve the criterion without having their
performance impaired by fatigue.

It will be evident from the figures above
that MLAT V makes it possible to estimate the total
time an individual will require to achieve a
professionally useful standard of performance. If
a student does not have the requisite amount of time

98

to spare in his schedule, he can be advised not to
commence the language study. Failure in this field
is often the product of bad planning and this can
not be avoided.

The average score for natural science under-
graduates was 16. For others it was 18. The
difference is probably due to the fact that most
of the scientists were men and a large proportion
of the others were women. Carroll and Sapon also
noted a marked difference between the performance
of the sexes. The figures in the present case are,
however, of limited interest because the non-
scientists had more language learning experience
and this may have had some influence on the scores.
In general the test was well designed to minimize
such factors. It begins with data presented as if
for a recall task, which turns out to be a multiple
choice problem, and this is likely to reduce the
advantages of previously learnt techniques.

BIBLIOGRAPHY

1. CARROLL, J.B. (1958) 'A Factor Analysis of two Foreign Language Aptitude Batteries', Journal of General Psychology 59 (3-19).

2. CARROLL, J.B. (1959) 'Use of MLAT in Secondary Schools', 16th Yearbook NCMUE (155-159).

3. CARROLL, J.B. (1961) The Study of Language, Harvard University Press.

4. CARROLL, J.B. (1962) 'The Prediction of Success in Intensive Foreign Language Training', in Training Research and Education (ed. R. Glazier) University of Pittsburgh Press (87-136)

5. CARROLL, J.B. and SAPON, S.M. (1959) MLAT Manual, Psychology Corporation.

6. HALSALL, E. (1969) 'Linguistic Aptitude', Modern Languages Vol. L. No.1. (18-23).

7. HARDING, F.D. Jr. (1958) 'Tests as Selectors of Language Students', Modern Language Journal 42, (120-122).

8. LADO, R. (1961) Language Testing, Longmans.

9. LOVELL, K. (1958) Educational Psychology and Children, University of London Press.

10. MACKEY, W.F. (1965) Language Teaching Analysis, Longmans.

11. PIMSLEUR, P. (1964) 'Testing Foreign Language Learning', Trends in Language Teaching (A. Voldman), McGraw-Hill (1965) (175-186).

12. PIMSLEUR, P., STOCKWELL, R.P., and COMLEY, A.L. (1962) 'Foreign Language Learning Ability' Journal of Educational Psychology 53. No.1. (15-26).

13. SAPON, S.M. (1955) 'A work sample test for Foreign Language Prognosis', Journal of Psychology 39, (97-104).

106

100

14. WHITE, N.   (1968-69)   Lectures in Psychology
     and Language Learning, Language Centre,
     University of Essex.

15. COX, R.   (1966)   Notes on the Use of a
     Reasoning Test, University of Essex mimeo-
     graph.

16. PARKIN, M. and HOMEWOOD, G. (1969) Admission
     Procedures: Research Report, 1969,
     University of Essex mimeograph.

17. HLYWOOD, J. (1969)   'Withdrawal from Higher
     Education', Student Wastage, paper prepared
     for the AUT Working Party on Productivity,
     Nov. 1969, University of Lancaster mimeograph.

18. DAVIES, A.   (1968)   Language Aptitude and
     Proficiency Project, Report to Steering
     Committee June 1968.

19. WATTS, W.J.   (1970)   Reception and Feedback in
     Audition, University of Sussex mimeograph,
     Jan. 1970.

20. WATTS, W.J.   (1969)   Some Effects of Hearing
     Impairment on a Small Student Population,
   University of Sussex mimeograph, Oct. 1969.

107

APPENDIX

Course Mark Analysis of Variance

	SS	DF	MS	VR	SL
Between Languages	9686·0	3	3227·7	19·78	0·1%
Between Courses	536·2	1	536·2	3·29	NS
PLS Regression	6161·4	4	1540·3	9·44	0·1%
MLAT Regression	1848·8	5	369·8	2·27	5%
Residual Error	79491·1	487	163·2		
Total	97723·5	500	195·4		

Exam Mark Analysis of Variance

	SS	DF	MS	VR	SL
Between Languages	8056·3	3	2685·4	24·90	0·1%
Between Courses	64·5	1	64·5	0·60	NS
PLS Regression	15744·8	4	3936·2	36·50	0·1%
MLAT Regression	3548·1	5	709·4	6·58	0·1%
Residual Error	52521·8	487	107·8		
Total	79935·5	500	159·9		

Correlations

Comark $\boxed{0·433}$

R $\boxed{0·587}$
Exmark $\boxed{0·586}$

Cos^{-1} 0·6183 Cos^{-1} 0·9991
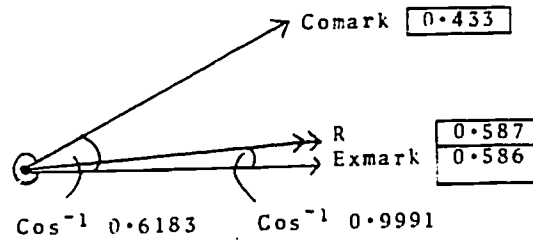
MLAT Variables on Predictor Variables

Demonstrating the extent to which MLAT results depend on previous language study:-

	a	n_A	n_O	n_C	Constant	Regression SS	Total SS
m_T	2.96	13.99	11.10	5.17	89.20	65,432	255,940
	1.64	1.27	1.49	1.53			
m_S	1.74	9.72	7.36	3.15	41.20	30,592	114,093
	1.08	0.84	0.77	1.01			
m_1	0.79	2.97	2.82	1.56	26.02	3,397	79,015
	0.71	0.55	0.51	0.66			
m_2	0.43	1.30	0.92	0.46	21.96	539.9	5764.7
	0.27	0.21	0.20	0.25			
m_3	0.62	2.75	1.88	1.02	13.82	2,271	33,848
	0.67	0.52	0.48	0.62			
m_4	0.77	4.33	3.00	0.47	16.17	6,226	21,877
	0.47	0.36	0.34	0.44			
m_5	-	2.66	2.49	1.66	12.75	2,641	17,313
		0.35	0.33	0.42			

Canonical Analysis:-

$$0.002\ m_1 + 0.018\ m_2 + 0.018\ m_3 + 0.112\ m_4 + 0.052\ m_5$$
$$\sim 0.22a + 1.23\ n_A + 0.92\ n_O + 0.30\ n_C$$

109