DOCUMENT RESUME

ED 050 136 TM 000 521

AUTHOR	Ilks, A. I.
TITLE	The Modern Language Aptitude Test in a Peace Corps Context: Validity, Expectancies, and Implications for Further Research.
SPONS AGENCY	Peace Corps (Dept. of State), Washington, D.C.
REPORT NO	I-68-03
PUB DATE	Jun 68
NOTE	24p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*Aptitude Tests, Expectancy Tables, Intensive Language Courses, Language Instruction, *Language Proficiency, Language Research, *Language Tests, *Predictive Ability (Testing), Predictive Validity, Predictor Variables, *Uncommonly Taught Languages
IDENTIFIERS	MLAT, *Modern Language Aptitude Test, Peace Corps

ABSTRACT

The language training program for Peace Corps participants is investigated. The objectives were: to investigate the validity of the Modern Language Aptitude Test (MLAT) for predicting end-of-training Foreign Service Institute (FSI) language proficiency; to discover the relative difficulty of various target languages; to develop tentative expectancy tables to enable prediction of a trainee's likely proficiency at end of training; and, to outline the need for and nature of an ongoing program of Peace Corps language research. Target language, total instructional hours, MLAT scores, initial and final FSI proficiency ratings were the variables considered. Findings indicated that MLAT has modest, though statistically reliable, predictive power for all languages; greater validity for common than for exotic languages; greater predictive power for shorter instructional periods (less than 280 hours) if language is disregarded; and taking language type and training time into consideration, greatest predictive power in common language projects involving 280 or more hours of training. Systematic proficiency differences were observed among languages, implying differences in relative difficulty for American students. More vigorous and innovative research is recommended to identify factors of importance in language training. Some thirty languages are grouped and listed in order of difficulty, and expectancy tables indicate the likelihood of attaining various levels of end-of-training proficiency. (Author/LR)

ED050136

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

THE MODERN LANGUAGE APTITUDE TEST IN A PEACE CORPS CONTEXT: VALIDITY, EXPECTANCIES, AND IMPLICATIONS FOR FURTHER RESEARCH

June 1968

Research Report P-68-03

Principal Investigator

A. I. Fiks

Peace Corps Contract

PC 80-1535

000 521

Staff

Barbara A. Zattiero
Frances H. Muth

Alexandria, Virginia

SUMMARY AND RECOMMENDATIONS

The research described in this report had four goals:

1. To investigate the validity of the Modern Language Aptitude Test (MLAT) for predicting end-of-training FSI language proficiency in Peace Corps trainees,
2. To discover the relative difficulty of various target languages as inferred from the data.
3. To develop tentative expectancy tables which would permit prediction of a trainee's probable language proficiency at the end of training.
4. To outline the needs for and the nature of an ongoing program of Peace Corps language research.

Procedure

The trainee data was reported to us on PC Form 1004. Our sample consisted of data on 1,731 trainees in sixty-four Summer/Fall 1967 training projects. The variables considered were: target language, total hours of instruction, MLAT, initial FSI proficiency rating, and final FSI proficiency rating. The outline for an ongoing program of language research is based in part on Language Training Documentation (LATRAD) inputs.

Findings

The following capsule statements may be made, on the basis of our data, although the reader should consult the body of the report for fuller treatment of the provisos and ramifications.

MLAT Validity

1. The MLAT has modest, though statistically reliable, predictive power for the criterion of end-of-training FSI rating in common as well as exotic languages.
2. MLAT validity is somewhat greater for common than for exotic languages.
3. Disregarding language, MLAT has more predictive power in shorter periods of language instruction (under 280 hours) than in longer ones.
4. Taking language type and training time into consideration, MLAT's predictive power is greatest in common language projects involving 280 hours or more of training.

Proficiency Differences Among Languages

5. Systematic differences in average proficiency among languages exist, implying, in part, differences in relative difficulty for American students.
6. Ewe and Swahili are near the top of the list (easiest); Thai and Korean near the bottom (hardest).

Language Proficiency Prediction

7. Expectancy charts are presented which indicate the likelihood of various FSI proficiency outcomes, given a trainee's MLAT, the target language, and the total number of instructional hours in the training project for which he is programmed.
8. Attainment of an S-2 rating, identified by Carroll (1966) as a minimum qualifying level for PC field effectiveness, is quite unlikely under present circumstances except:

a. in easier languages for high MLAT (> 58) trainees, with 280 hours or more of instruction; and/or

b. in French and Spanish for trainees with some initial proficiency.

9. In some cases, at least, MLAT and training time seem to compensate for each other, so that shorter training with higher MLAT permits about the same FSI prediction as longer training with lower MLAT.

10. For the more difficult languages, increasing training time over the range contained in our data does not appear to improve FSI proficiency expectations in any major way except, to some extent, for the lowest MLAT group.

Continuing In-House Language Research

Research on language training can serve three functions: documentation, quality control on training, and innovation. Documentation is needed because there is:

a. high turnover of language training personnel,

b. great variation in education and other background factors of language training personnel,

c. often short lead-in time for training projects.

Quality control on training is still largely in an anecdotal, impressionistic stage making evaluation of pedagogic efforts and approaches difficult. More rigorous, innovative research may help to better isolate the actual factors of importance in language training.

Recommendations

The following suggestions seem warranted on the basis of the results.

1. Assuming a reasonable selection ratio (surplus of applicants over trainee positions) the MLAT is a meaningful screening device on which to base recommendations to training projects, especially those involving French, Spanish, or Portuguese.
2. The setting of training goals should take into consideration systematic differences in proficiency which are realistically achievable in different languages, reflecting, in part, their relative difficulty.
3. The expectancy tables included in this report may be used to predict various probable FSI proficiency outcomes as a function of: target language, total training time, and trainee's MLAT score.
4. With regard to further implications of these quantitative analyses, additional data should be collected in order to extend the coverage to more languages (than the thirty-one covered here); and in order to allow greater confidence in predictions made from the expectancy tables.
5. One such possible implication that requires further data, is the utility/non-utility of extending the total number of hours of language training in the more difficult languages.
6. With regard to continuing language research:
   a. Evaluation/documentation of training activities carried out by contractor organizations should be conducted in-house by PC/Washington research and language training personnel.
   b. Evaluation/documentation of training activities carried on in-house (e.g., Virgin Islands Training Center, Leland, Escondido, Puerto Rico) should be conducted by outside research organizations under contract.

c. One of the Peace Corps training sites should be designated as a Training Laboratory where rigorous research of an experimental nature can more adequately be carried out.

CONTENTS

Page

TABLES

INTRODUCTION

Many questions of language training do not lend themselves to very strict scientific analysis. Several basic points, however, do permit at least a systematic/quantitative research approach. The validity of the Modern Language Aptitude Test (MLAT) for Peace Corps with end-of-training Foreign Service Institute (FSI) language proficiency ratings as the criterion is one such point. Whether the predictive power of the MLAT is constant or varies as between different languages and different training durations, is another. The relative difficulty of various target languages for PC trainees is an issue toward which hard data would at least permit an inference to be made.

Based on the foregoing, to the extent that language aptitude, training time, and target language difficulty are related to achievement, it should be possible to construct some expectancy tables. Such charts would permit prediction of probable end-of-training FSI proficiency given the relevant antecedent information. It is such questions that the quantitative research reported herein is intended to answer.

In a more procedural vein, we outline in this report some thoughts on the needs for a continuing language research program to be performed in-house and/or through contractors.

PROCEDURE

The data source for the quantitative analyses in this report and the conclusions they permit are PC Form 1004. This form reports: target language, hours of instruction, Modern Language Aptitude Test (MLAT) score,[1] initial (start of training) FSI Speaking Proficiency rating, and final (end of training) FSI rating for each trainee.

The sample tapped in this study consists of trainee data from sixty-four Summer/Fall 1967 training projects. The total number of trainees involved was 1,731 although some analyses were based on fewer cases. [2]

The outline of a suggested language training research program is based in part on Language Training Documentation (LATRAD) inputs and in part on visits to training sites.

1/ The scores reported and used throughout this report are the PC MLAT which are MLAT raw scores (for Parts 3, 4, and 5) standardized with mean = 50 and standard deviation = 10.

2/ The problem of missing data was present to a considerable extent. We have no way of determining to what degree this may be biasing to the findings.

FINDINGS

In this section we present data concerning three substantive areas of interest: 1) the utility of the MLAT in Peace Corps selection/placement/training; 2) systematic proficiency differences achieved in training programs involving different languages; and 3) the expectancy of certain levels of language proficiency that it is reasonable to have for a trainee given a certain target language, his language aptitude, and the duration of his training.

MLAT Validity

Past research has established that the MLAT has considerable validity for predicting end-of-training criteria in all sorts of student groups from ninth-grade school children to military personnel at the Army Language School. (Carroll and Sapon, 1959.)

In a Peace Corps population, Carroll (1966) has reported validity coefficients of .32 and .41 for two subgroups of Spanish language students. The criterion in that case was a 5-point rating given each trainee by the language staff at the end of training. Against a more objective criterion of language proficiency, the MLA Cooperative test, the validity coefficient for the MLAT was .67 and .33 for the same respective subgroups.

In the current research we use the actual FSI Speaking proficiency rating as the criterion. The FSI rating is, after all, the most widely known and clearly recognized language proficiency index. The data in Table 1 can be used to answer several questions:

1. Is MLAT predictive of FSI S-rating for Peace Corps trainees generally, i.e., in languages other than Spanish?

2. Is MLAT differentially predictive as between common and exotic languages?

3. Is MLAT differentially predictive in shorter or longer training periods?

4. Taking both language type and total hours of training into consideration, in which of the possible situations is MLAT most predictive?

5. What are the implications of these findings to Peace Corps utilization of MLAT?

In answer to the first question, the first two breakdowns of Table 1 are relevant. Note that for common as well as exotic languages, there was a systematic decrease in average MLAT as one descends from trainees who achieved S-3 level or better to those who reached S-2+, S-2, etc. The only reversals in the regularity occurred for the lowest proficiency group in each case. The correlation coefficients which characterize this regularity are r = .30 and .11,

ERIC

respectively. Though quite modest, these correlations are significantly greater than zero. Thus, the question must be answered in the affirmative on the basis of these data: MLAT is moderately predictive of end-of-training language proficiency in Peace Corps training generally in common as well as exotic languages.

With regard to the second question posed above, it appears on the basis of these data that the MLAT is a more salient predictor in the common languages (i.e., French, Spanish, Portuguese) than it is for exotic languages. The quantitative difference being considered in this analysis is that between r = .30 for common languages and r = .11 for exotic languages. That particular difference based on the present number of cases is extremely unlikely to have occurred merely by chance ($p < .002$).

Similarly, the third and fourth breakdowns of Table 1 compare the predictive power of the MLAT for shorter vs. longer language training programs. Disregarding language type then, it appears that the MLAT is somewhat more predictive in shorter programs (under 280 hours) where r = .21 than it is for programs with larger numbers of language hours, where the correlation is only .08. Although this absolute difference in correlation coefficients is not dramatic in size, it is a statistically stable finding ($p < .02$). 3/

To consider the fourth question above, the data were analyzed so as to consider language type and total hours of instruction simultaneously. These results are shown in the last four breakdowns of Table 1. Note that the highest correlation is found for common languages of 280 or more hours of instruction, where r = .44. There is a seeming disparity between this statement and the previous finding that generally, disregarding language type, the MLAT is more predictive in shorter programs. The latter analysis where both factors are taken into consideration, i.e., language type and training duration, would seem to be the more useful.

With shorter training programs, i.e., breakdowns 5 and 7 of Table 1, there is no stable difference between the predictive power of the MLAT (i.e., where r = .18 and .21, respectively). However, for longer projects, i.e., breakdowns 6 and 8 of Table 1, the MLAT appears to have a good deal more predictive power for common languages than for exotic ones where the correlation drops to zero.

The answer to the fifth question, implications of these findings for Peace Corps training and selection, seems to be the following. The modest but statistically stable correlations between the MLAT and FSI Speaking proficiency rating at the end of training suggest that the MLAT may properly be taken into consideration in inviting applicants to training projects on the premise that foreign language achievement is an important goal of training. The modest size of the coefficients, however, does not permit use of the test for purposes of individual counseling. With regard to placement it would appear that an applicant's MLAT should carry somewhat heavier weight if he is applying for a common language program than if he is aiming for an exotic language program, since the predictive power of the MLAT is lower in the second case. If, however, MLAT scores are not available to cognizant PC personnel who issue invitations to training and/or to language coordinators at the very beginning of training, this information obviously cannot be used either for trainee selection or to form homogeneous language classes, and becomes rather academic.

3/ These correlations between MLAT and FSI rating are roughly comparable in size to those found between MLAT and Selection Board Final Rating by Krug (1962).

Table 1

THE RELATIONSHIP BETWEEN APTITUDE (MLAT) AND PROFICIENCY (FSI RATING) FOR COMMON AND EXOTIC LANGUAGES AS WELL AS SHORTER AND LONGER TRAINING DURATIONS

Language Type	Total Hours of Instruction	No. of Trainees	Mean MLAT	Final FSI Speaking Proficiency	MLAT - FSI Correlation Coefficient [a]
Common	All	320			
		6	62.33	S-3 or higher	
		22	60.22	S-2+	
		73	58.02	S-2	
		158	55.34	S-1+	r = .30
		46	52.58	S-1	
		15	54.19	S-0+ or lower	
Exotic	All	1051			
		15	61.13	S-3 or higher	
		32	58.31	S-2+	
		153	58.13	S-2	
		344	55.55	S-1+	r = .11
		321	55.27	S-1	
		186	55.88	S-0+ or lower	
All	0-279	843			
		9	62.21	S-3 or higher	
		29	58.51	S-2+	
		130	57.89	S-2	
		343	55 51	S-1+	r = .21
		247	54.42	S-1	
		85	53.30	S-0+ or lower	
All	280 or more	474			
		12	60.91	S-3 or higher	
		25	59.75	S-2+	
		95	58.39	S-2	
		159	55.44	S-1+	r = .08
		104	55.70	S-1	
		79	58.32	S-0+ or lower	

[a] Serial correlation was calculated (Jaspen, 1946) to accomodate FSI ratings, which constitute a segmented variable and which therefore present certain difficulties for the usual Pearson product-moment correlation.

Table 1 (continued)

THE RELATIONSHIP BETWEEN APTITUDE (MLAT) AND PROFICIENCY (FSI RATING) FOR COMMON AND EXOTIC LANGUAGES AS WELL AS SHORTER AND LONGER TRAINING DURATIONS

Language Type	Total Hours of Instruction	No. of Trainees	Mean MLAT	Final FSI Speaking Proficiency	MLAT - FSI Correlation Coefficient [a]
Common	0-279	157			
		0	...	S-3 or higher	
		12	60.16	S-2+	
		28	57.64	S-2	
		78	55.30	S-1+	r = .18
		26	54.84	S-1	
		13	55.30	S-0+ or lower	
Common	280 or more	163			
		6	62.33	S-3 or higher	
		10	60.29	S-2+	
		45	58.26	S-2	
		80	55.38	S-1+	r = .44
		20	49.64	S-1	
		2	47.00	S-0+ or lower	
Exotic	0-279	686			
		9	62.21	S-3 or higher	
		17	57.35	S-2+	
		102	57.97	S-2	
		265	55.57	S-1+	r = .21
		221	54.37	S-1	
		72	52.94	S-0+ or lower	
Exotic	280 or more	311			
		6	59.49	S-3 or higher	
		15	59.39	S-2+	
		50	58.51	S-2	
		79	55.50	S-1+	r = .00
		84	57.14	S-1	
		77	58.62	S-0+ or lower	

[a] Serial correlation was calculated (Jaspen, 1946) to accomodate FSI ratings, which constitute a segmented variable and which therefore present certain difficulties for the usual Pearson product-moment correlation.

In evaluating the above validity coefficients it must be remembered that one possible reason for their modest size is a psychometric weakness in the criterion, i.e., the FSI ratings. The reliability of those ratings themselves limit the validity which a predictor test can demonstrate. Secondly, it should also be borne in mind that the MLAT scores in question are rather high, thus representing a somewhat narrow range of aptitude. This may be deduced from the fact that the MLAT scores, as given, are generated PC MLAT s which were standardized with a mean of 50 and a standard deviation of 10 on early groups of volunteers.

Proficiency Differences Among Languages

The substantive question to which we would like to address ourselves here is: Are there any reliable differences in difficulty among the various languages studied by PC trainees, eliminating any extraneous variables? Unfortunately, the data do not allow us to answer this question. The reason for this is that many extraneous factors are varying from project to project besides the fundamental one of which language is being studied. In Table 2 we have eliminated statistically one of these extraneous variables, namely, differences in duration of training. Another, differences in initial proficiency, has been controlled by considering only trainees with zero starting ability in any of the languages. Other factors however, like quality of instructor personnel, adequacy of text materials, motivation of trainees cannot so easily be controlled for. Therefore, the question to which the data of Table 2 really address themselves is: Are there any reliable differences in final FSI Speaking proficiency ratings among groups of PC trainees studying some thirty-one different languages?

In Table 2 we have arranged the languages in rank order of average proficiency achieved. Eliminating initial proficiency and amount of training as factors, we note that Ewe produced the highest average proficiency; Korean, the lowest. Another assumption in making this kind of comparative statement with regard to languages is that the standards of FSI testing are completely constant across languages, a premise not impervious to question. The analysis of covariance on these data results in an F ratio of 29.9 which allows the statement at a very high level of statistical confidence ($p < .001$) that, holding training time and initial proficiency constant, the average FSI proficiency among these thirty-one languages is not the same.

This is not to imply, however, that neighboring languages on the list like Turkish, for example, (mean adjusted final FSI score of 1.61) yields reliably higher average proficiency than Malay (1.59). As indicated by the approximate standard errors of the means given in the table, the neighboring means fall well within a one standard error band around any comparison mean. Thus the conclusion that the overall F is significant so that overall differences in average FSI proficiency exist, cannot necessarily be applied to any two languages near each other.

The break in the table between Punjabi and Visayan is arbitrarily placed there to divide the languages into two groups (A and B) with roughly equal numbers of trainees. This grouping will be used in a subsequent analysis.

Table 2

AVERAGE FINAL FSI SPEAKING PROFICIENCY RATING FOR VARIOUS LANGUAGES ELIMINATING EFFECTS OF DIFFERENT AMOUNTS OF TRAINING AND ANY INITIAL PROFICIENCY (N = 1731)

Group	Language	Region	No. of Trainees	Adjusted Mean Final FSI Score a/	Standard Error of the Mean b/
	Ewe	Africa	13	2.02	1.22
	Swahili	Africa	8	1.91	1.47
	Ilocano	East Asia Pacific	28	1.81	1.09
	Wolof	Africa	3	1.79	1.67
	Twi	Africa	35	1.69	0.68
	Tumbuka	Africa	7	1.68	2.14
	Ilocano	East Asia Pacific	70	1.68	0.53
	Nyanja	Africa	28	1.68	0.95
	Hausa	Africa	12	1.66	0.90
A	Turkish	NANESA	107	1.61	0.89
	Malay	East Asia Pacific	139	1.59	0.32
	Sotho	Africa	51	1.59	0.52
	Western Arabic	NANESA	43	1.57	0.87
	Portuguese	Latin America	36	1.55	1.08
	Persian, Afghan	NANESA	52	1.52	0.68
	Tagalog	East Asia Pacific	159	1.52	0.33
	Persian	NANESA	24	1.43	1.41
	Spanish	Latin America	390	1.41	0.25
	French	Africa, NANESA	33	1.41	0.76
	Punjab	NANESA	7	1.37	0.71
	Visayan	East Asia Pacific	112	1.30	0.47
	Pashto	NANESA	26	1.23	1.05
	Telugu	NANESA	54	1.21	0.40
	Nubian	NANESA	15	1.18	0.83
	Fula	Africa	2	1.13	0.00
B	Nepali	NANESA	32	1.00	0.59
	Mandinka	Africa	12	0.96	0.94
	Hindi	NANESA	11	0.94	1.35
	Amharic	Africa	76	0.94	0.30
	Thai	East Asia Pacific	37	0.56	0.51
	Korean	East Asia Pacific	109	0.35	0.37
			1,731		

a/ The means have been adjusted through an analysis of covariance technique to eliminate the effects of differential training time. The scale is parallel to the FSI dimension; 2.00 = S-2, 1.50 = S-1+, 1.25 = halfway between S-1 and S-1+, etc.

b/ Approximate values due to computation from unadjusted scores.

A comparison between Table 2 and FSI breakdown of language difficulty matches fairly well, at least in some particulars. Swahili, for example, is in Group 1 (easy) for FSI (along with French, Spanish and Portuguese); and it is near the top of the list in Table 2. Malay, Persian and Turkish are in the middle range both in FSI and in the list of Table 2. Korean is classified as the most difficult in both lists. The two lists are not completely comparable, however, particularly in that Table 2 is much longer than that in use by FSI. (1966)

Language Proficiency Prediction

Having established that the Modern Language Aptitude Test has modest predictive power and that systematic proficiency differences appear to exist among various languages even when holding training duration constant, it now becomes meaningful to ask the following questions.

1. Given a PC trainee whose MLAT score falls within a certain range and who is being assigned to study language X in which he will receive Y number of hours of instruction, what is a reasonable expectation of his proficiency at the end of training?

2. How likely is it that he will attain the S-2 level?

3. Do the two factors of language aptitude and duration of training compensate for each other, i.e., does longer training time make up for lower MLAT?

4. In what way is the expectation altered if trainees have some initial proficiency in the target language?

Tables 3 and 4 deal respectively with language Groups A and B as identified in Table 2. Table 3 presents data for language Group A which contains languages that are easier or (at least) in which higher final FSI scores occur. Language Group B, in Table 4, contains languages which appear to be harder/result in lower average FSI proficiency.

We will first attempt to answer questions 1, 2, and 3 above for language Group A as indicated in Table 3. An expectancy of 1.000 will represent absolute certainty and .000 will mean absolute impossibility. Thus, for a random trainee exposed to less than 280 hours of instruction in a language in Group A, the chances of achieving no more than S-1 proficiency, i.e., S-1 or less, is .337 if his MLAT is under 52; remains at about one-third (.325) if his MLAT is between 52 to 58; but drops to .214 if his MLAT is over 58. Note that the parallel expectancies for the same individual to reach no higher than S-1 if he were to get 280 hours or more of language training are thereby considerably lowered. Table 3 may be used to state the likelihood of any other final FSI proficiency outcome in a similar manner.

In particular, since Carroll (1966) identified the S-2 level as a minimum qualifying benchmark for effectiveness in the field, it becomes important to answer the second question above, namely what is the likelihood that an

Table 3

RELATIONSHIP BETWEEN APTITUDE (MLAT), TOTAL HOURS OF LANGUAGE INSTRUCTION AND FINAL FSI SPEAKING PROFICIENCY RATING IN LANGUAGE GROUP A FOR TRAINEES WITHOUT INITIAL PROFICIENCY (N = 946)

MLAT Range	Final FSI Speaking Proficiency Rating	Less than 280 hours	280 Hours or More
		N = 190	N = 83
Under 52	S-1 or less	33.7 %	27.7 %
	S-1+	50.5	48 2
	S-2	12.6	20.5
	S-2+ or more	3.2	3.6
		N = 200	N = 116
52 - 58	S-1 or less	32.5 %	12.9 %
	S-1+	43.0	52.6
	S-2	18.5	23.3
	S-2+ or more	6.0	11.2
		N = 243	N = 114
Over 58	S-1 or less	21.4 %	3.5 %
	S-1+	44.4	28.1
	S-2	25.5	46 5
	S-2+ or more	8 6	21.9

Table 4

RELATIONSHIP BETWEEN APTITUDE (MLAT), TOTAL HOURS OF LANGUAGE INSTRUCTION AND FINAL FSI SPEAKING PROFICIENCY RATING IN LANGUAGE GROUP B FOR TRAINEES WITHOUT INITIAL PROFICIENCY (N = 375)

MLAT Range	Final FSI Speaking Proficiency Rating	Less than 280 Hours	280 Hours or More
		N = 63	N = 23
Under 52	S-1 or less	80.9 %	73.9 %
	S-1+	17.5	21.7
	S-2	1.6	4.3
	S-2+ or more	0.0	0.0
		N = 66	N = 62
52 - 58	S-1 or less	74.2 %	77.4 %
	S-1+	25.8	21.0
	S-2	0.0	1.6
	S-2+ or more	0.0	0.0
		N = 61	N = 100
Over 58	S-1 or less	68.8 %	79.0 %
	S-1+	24.6	19.0
	S-2	6.6	2.0
	S-2+ or more	0.0	0.0

individual of given aptitude will reach this minimum qualifying level? If a trainee's MLAT is under 52 and he gets less than 280 hours of training, his chances of achieving at least an S-2 level are .158 (i.e., 12.6% plus 3.2%). If that same trainee get 280 hours or more, his probability for "success" increases to .241. In a parallel fashion if his MLAT is between 52 and 58, his likelihood of achieving the S-2 level is .245 with less than 280 hours and .345 with longer instruction. Students high in language aptitude (over 58) increase their S-2 likelihoods markedly to .341 and .684, respectively.

To the third question, regarding the possible compensatory nature of aptitude and training time, we note from Table 3 that, at least for Group A languages, the answer must be in the affirmative. Note that the expectancy for a low MLAT trainee (under 52) with long training time is .241 (20.5% and 3.6%) of reaching S-2 level which is very close to the medium aptitude trainee (52 to 58) with less than 280 hours for whom the likelihood is .245. Similarly the medium aptitude trainee with long training has a likelihood of .345 of reaching the S-2 level, the high aptitude (over 58) trainee has a very similar .341 chance with the shorter training duration.

The general pattern of results in Table 4 is similar to that in Table 3 with one major exception. By definition in Table 4 the performance range on the FSI is considerably lower than that for languages in Group A. Thus, for Group A languages, the typical/modal/most likely language training outcome regardless of MLAT or training duration was, with one exception, S-1+. The one exception in Table 3 was for high aptitude, long training time outcome where the modal category was S-2. Turning to Table 4, however, we find that in all breakdowns the most likely outcome (i.e., the largest per cent) is S-1 or less.

One other respect in which the data for language Group B appear to differ from those in language Group A is as follows. For language Group B increasing training time to 280 hours or more does not appear to improve the FSI proficiency expectations in any meaningful way. The possible exception is in the low MLAT group where the likelihood of reaching at least an S-1+ level is .260 with 280 hours or more compared to .191 for the shorter time. It may also be seen in Table 4 that for practical purposes the likelihood of reaching the Carroll identified minimum qualifying level of S-2 ranges between zero and .066, i.e., is highly unlikely to occur by the end of the training period.

In Table 5 we consider the likelihood of various expectations of end-of-training proficiency for trainees who enter the program with some initial ability. These data are for French and Spanish only. It may be seen in Table 5 that the most likely outcome for these trainees is the S-2 rating.

The likelihood that a trainee with some initial proficiency will reach at least the S-2 level in French or Spanish with less than 280 hours is .680 (i.e., 31.9% plus 22.2% plus 4.2% plus 9.7%) if his MLAT is under 52, .756 if his MLAT is 52 to 58 and .794 if his MLAT is over 58. Similarly, with 280 hours or more

of training, the likelihood of his reaching the S-2 level is .685, .760 and .880, respectively. In other words, with some prior skill, the S-2 level is a realistic training goal in French and Spanish language programs.

The foregoing expectancy charts, i.e., Tables 3, 4, and 5 should be augmented and updated with additional trainee data in order to increase the confidence one can place in the expectancy statements or predictions.

Table 5

RELATIONSHIP BETWEEN APTITUDE (MLAT), TOTAL HOURS OF LANGUAGE INSTRUCTION AND FINAL FSI SPEAKING PROFICIENCY RATING IN FRENCH AND SPANISH FOR TRAINEES WITH SOME INITIAL PROFICIENCY (N = 522)

MLAT Range	Final FSI Speaking Proficiency Rating	Less than 280 hours	280 Hours or More
		N = 72	N = 38
Under 52	S-1 or less	5.5 %	2.6 %
	S-1+	26.4	28.9
	S-2	31.9	34.2
	S-2+	22.2	5.3
	S-3	4.2	15.8
	S-3+ or more	9.7	13.2
		N = 119	N = 46
52 - 58	S-1 or less	6.7 %	2.2 %
	S-1+	17.6	21.7
	S-2	31.9	34.8
	S-2+	16.8	21.7
	S-3	15.1	13.0
	S-3+ or more	11.8	6.5
		N = 180	N = 67
Over 58	S-1 or less	2.8 %	0.0 %
	S-1+	17.8	11.9
	S-2	33.?	35.8
	S-2+	24.4	20.9
	S-3	13.3	14.9
	S-3+ or more	7.8	16.4

CONTINUING IN-HOUSE LANGUAGE RESEARCH

Training in foreign languages is a principal component of the Peace Corps training system. The need for continuing research on language training has three aspects to it: documentation, quality control on training, and innovation.

Documentation

Systematic, descriptive documentation of training philosophy, operational procedures, and materials is a needed research function to overcome the data gap which makes the planning of future language training difficult and tenuous.

This is particularly a problem due to three factors previously described (Fiks, 1967):

1. There is a high turnover of language training personnel. In effect this removes the possibility of a pyramidal trial and error approach to increasingly effective methods that would be possible with a more stable staff.

2. There is a great variation in education and other background of Peace Corps language personnel. This condition is true for language coordinators as well as for instructor personnel. The result is considerable unevenness in the caliber of training projects with the more effective ones largely hinging on the individual personality and approach of the particular coordinator/teachers rather than on systematic training policy or practices.

3. There is often short lead-in time with ineffective communications during the project planning stage. This condition makes it highly desirable to have documentation available, organized, and stored in such a way as to make quick retrieval and dissemination to new personnel feasible. Procedures for data collection, coding and storage and retrieval are contained in Fiks (1967).

Quality Control on Training

The evaluation of ongoing and past language training activities is still largely in the anecdotal, impressionistic stage. More systematic quality control requires rigorously documented accounts of training procedures on the one hand, and hard language proficiency measures and field criterion data on the other.

As a general rule, we would suggest that such documentation/training quality control be done in-house by Peace Corps research personnel in conjunction with PC/Washington language training personnel for those projects which have been contracted out to universities and other organizations. Conversely, for language

training taking place within PC training sites such as Virgin Islands, Escondido, Leland we would suggest the contracting of any language research to be done to outside organizations.

These suggestions are made based on personal experience and the difficulty which exists in coordinating data collection at contract training sites on the one hand, and the danger of non-objectivity of in-house self-evaluation on the other.

Innovation

Innovative research on language training is needed basically because for the most part, trainees fail to achieve the empirically determined minimum qualifying level by the end of training (Carroll, 1966). This is clearly seen in our Tables 3 and 4 above. The fact that it is at least theoretically possible to reach the S-2 level is established in Table 3 where among trainees without prior proficiency, with MLAT below 52, and with less than 280 hours of instruction, 15.8% do achieve the S-2 level. Assuming valid and reliable proficiency ratings for the moment, this result must be due to some trainee characteristics such as motivation and/or some training system characteristics, like having a particular type of teacher, a particular type of text, with particular types of environmental supports. Innovative research would help isolate the factors which are of importance to the training goal.

A secondary reason for a program of innovative research is that Peace Corps bears a certain obligation to contribute to the fund of knowledge about training technology. Having the resources, i.e., the subjects, the training sites, and the research resources, it behooves Peace Corps not only to draw upon the "state of the art" but to contribute to this growing science. We would urge that any such research be undertaken not on a post facto evaluation basis but be regarded and staffed and planned as true experiments. The question of systematic sampling and random assignment of trainees to experimental conditions requires particular attention.

In essence, what is being suggested here is that Peace Corps designate one or more of its training sites to include the function of training laboratory. It is only in this way that rigorous answers can be hoped for to questions such as the following:

1. What differential language proficiency effects are found as a result of hyper-intensive language training (HILT) at the beginning, versus the middle, versus the end of the training period?

2. How does 300 hours of HILT training compare with 300 hours of less densely packed training?

3. How can programmed instruction best be harnessed to PC training needs?

4. At what stage in the learning process is there the greatest payoff from the environmental support and reinforcement implied by the practice of "immersion"?

REFERENCES

Carroll, J. B., A Parametric Study of Language Training in the Peace Corps, Final Report, Peace Corps Contract PC-(W)-226, 1966.

Carroll, J. B. and S. M. Sapon, Modern Language Aptitude Test, Manual, The Psychological Corporation, New York, 1959.

Fiks, A. I., The Language Training Documentation System, Final Report, Peace Corps Contract No. 80-1524, August 1967.

Foreign Services Institute, Expected Speaking Achievement in Language Training, (mimeographed paper, 1966).

Jaspen, N., Serial Correlation, Psychometrika, 1946, II, pp. 23-30.

Krug, R. E., An Analysis of Eighteen Peace Corps Projects, Part I of the Final Report, Peace Corps Contract PC-(W)-59, October 1962.