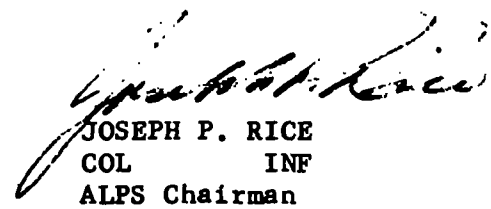
ARMY LINGUIST PERSONNEL STUDY (ALPS)

VOLUME II

DEFENSE LANGUAGE INSTITUTE (DLI)

BACKGROUND DATA AND INFORMATION (PART 2)

Summary of material and policy statements submitted by DLI, August 1975, in support of ALPS.



JOSEPH P. RICE
COL INF
ALPS Chairman

DEPARTMENT OF THE ARMY
OFFICE OF THE DEPUTY CHIEF OF STAFF FOR PERSONNEL
WASHINGTON, D.C. 20310

SUBJECT: Army Linguist Personnel Study (ALPS)

1. The material in this volume has been furnished by Defense Language Institute (DLI) in support of the ALPS.

2. For ease in using this document in conjunction with the ALPS all pages have been numbered in sequence and shown: (1) etc.

JOSEPH P. RICE
COL INF
ALPS Chairman

POSITION PAPER AND INFORMATION

TABLE OF CONTENTS VOLUME II

DLIR 25 Jun 75

POSITION PAPER

DLP TESTING POLICIES (DLPT AND DLAB)

1. PURPOSE: The purpose of this position paper is to briefly outline Defense Language Program (DLP) policies now being formulated as they concern the development of the new Defense Language Aptitude Battery (DLAB) and a proposed change to the present Defense Language Proficiency Test (DLPT) system.

2. DISCUSSION:

a. Defense Language Aptitude Battery (DLAB).

(1) In the late 1960's a new theoretical approach for measurement of language-learning aptitude appeared which held promise of improved prediction of academic failure in language training and, therefore, improved cost-effectiveness by screening out low-aptitude applicants for language training. Some of the new concepts were developed by COL Kibbey Horne, then Commandant, DLIWC, and by Dr. Antoine Al-Haik, now Chief, Test, Evaluation and Research Branch of the Office of Research and Development. Several different tests were developed and, through analysis of trial data, the best have been combined into a new battery, now referred to as DLAB.

(2) The procedures involved massive amounts of data and extensive, sophisticated statistical analysis. For advice on certain statistical problems that arose, and special computer analysis of the data, DLI obtained the assistance of Educational Testing Service.

(3) One of the concepts that showed promise early in the project was differential prediction of aptitude by target language, a hypothesis that particularly interested COL Horne. For assistance in factor analysis of DLI's experimental test forms, certain commercial tests, and other work related to that hypothesis, DLI contracted with San Jose State University. (At that time, DLI's own computer capability was minimal.)

(4) However, from all analyses to date of the present form of DLAB, it provides better general prediction of language aptitude than any other known tests, commercial or other-

wise. Differential prediction, that is the ability to predict success in a specific language, has not proven feasible to date.

(5) At present, DLI is collecting the final normative data. Data on Army and Air Force populations are now being prepared for analysis by DLI's Computer Systems Division and printouts are expected in July.

(6) Because of growing interest in the DLAB, DLI's policy is to accelerate data collection and operational implementation. Contingent upon data collection, data analysis, preparation of answer forms and other administrative tasks, operational implementation is planned for March 1976.

b. Proposed change to DLPT system.

(1) For some time, DLI has become increasingly aware of needed changes to the present Defense Language Proficiency Test (DLPT) system. These tests are administered by DLI as part of end-of-course testing and by the individual Services to inventory and evaluate language-qualified personnel in a variety of military specialities.

(2) The problems with the testing system are both technical and managerial. Therefore, the changes to testing policies and plans proposed earlier this month to the four Services did address both the technical and managerial issues and suggested a phased program of implementation by FY78.

(3) The new tests will be somewhat longer than the present DLPTs, but will otherwise look very much like them. That is, there will be a taped-recorded listening comprehension part and a reading comprehension part. DLI expects them to have at least as high reliability as the present tests, and substantially improved validity.

(4) The change that would be most apparent to the Services will be in the way DLPT scores are derived and reported. The present procedure is to convert raw scores to skill categories, called "levels," ranging from level-5 (proficiency equal to that of a native speaker) down to level-0 (no proficiency), by means of a conversion chart. Thus, for example, all linguists whose raw scores on the listening comprehension part of the DLPT I are in the range 40 to 47 are now simply categorized as level-3 in listening ability.

Since few trainees achieve higher than level-3, and few are as low as level-1 at graduation, the present system does little more than divide them into two groups: level-2 and level-3. Under the proposed procedure, raw scores would be converted to a standard scale (T-scale). Listening comprehension skills of individual linguists would under this plan, be recorded as T-scores, such as 39, 41, 47, 71, etc., permitting far more precise identification of the relative degree of foreign-language comprehension possessed by each.

(5) In addition to this effort, DLI has requested, and Army Research Institute is now processing, requests for two research projects. These, however, will affect only long-range plans for improvement of the validity and reliability of DLI's testing systems.

DLIR-D-T 4 Aug 75

MEMORANDUM FOR: Deputy Director/Director

SUBJECT: Biweekly Status Report for Work Unit 0021-1
Aptitude Test Assessment (21 July - 1 August 1975)

1. The analysis of the DLAB proof-test data is now well underway and it appears that the experimental form will require only minor modifications before becoming the final form (some of the Biographical items will be deleted).

2. The following table presents the relative predictive validities of DLAT and DLAB on students tested between January and May 1975:

	DLAT				DLAB				Grades	
	N	r	x	s	N	r	x	s	x	s
Arabic	140	.21	30.0	8.0	153	.40	72.0	15.4	83.0	10.9
Czech	70	.50	29.8	11.0	85	.64	70.0	17.0	85.8	7.7
French	43	.31	24.5	8.8	86	.55	61.9	13.6	87.8	6.2
German	99	.23	29.2	9.0	106	.43	68.5	14.5	87.4	7.7
Korean	78	.47	29.9	9.7	92	.55	67.3	15.2	83.7	10.9
MAFAC	27	.40	33.4	8.9	31	.54	69.3	15.1	81.4	10.4
RAFAC	79	.33	34.2	8.9	86	.34	76.2	14.0	83.5	10.6
Russian Basic	73	.57	29.2	9.9	86	.68	67.0	15.9	82.0	12.9
Spanish	66	.50	25.8	7.9	83	.59	63.2	14.7	86.4	9.1
Thai	27	.39	27.4	10.0	51	.53	67.5	16.8	88.8	6.9
Vietnamese	37	.40	30.1	7.8	38	.43	67.4	15.7	84.1	9.7
Chinese-Mandarin	75	.24	31.1	9.2	85	.62	70.4	16.6	82.0	8.8
TOTAL		.36				.50				

Keys: N - Number tested (not the same for DLAT and DLAB because all records could not be matched)

r - Validity coefficient

x - mean or average

s - standard deviation

(41)

The table requires several notes of explanation. First, the important feature is the pattern of correlations, not their absolute magnitudes in any single language. In every case DLAB is superior to DLAT, but it does not necessarily follow that DLAB is, for example, less predictive for RAFAC than for Russian Basic. The DLAB mean score for RAFAC indicates that the students in that course are a highly select group to begin with which makes their relative performance more difficult to predict. If the range of ability in RAFAC was similar to that in other courses, the predictive validity would undoubtedly be higher. Another factor affecting the validity coefficients is the manner in which student grades were collected. They were not obtained at a single point in time which means that their usefulness as a criterion in any language is largely determined by the degree to which student performance is consistent across time in that language.

3. Considering the criterion and restriction of range problems (i.e. students are already selected on the basis of aptitude), the validity of DLAB appears to be excellent, and preparations for producing the final form and supporting materials will begin immediately.

CALVIN R. PETERSEN
Project Officer

DRAFT

Abstract

The purpose of this research was to develop a language aptitude battery for use by the Defense Language Institute (DLI) which offers training in over 50 foreign languages. Predictor and criterion measures were collected on DLI students in 11 languages. Items were selected and grouped through canonical factor analysis of an experimental item pool. Validity was established through prediction of grades. These validities were then compared with those of other tests and predictor variables on the same student sample. The results were that the new battery (DLAB) produced consistently higher validities than other predictors. Differential prediction by language was also investigated, but evidence that it can be achieved was not found in this study.

THE DEVELOPMENT OF THE DEFENSE LANGUAGE APTITUDE BATTERY (DLAB)[1]

Calvin R. Petersen and Antoine R. Al-Haik
Defense Language Institute

The Defense Language Aptitude Test (DLAT) was developed approximately twenty years ago as a device for the selection of male army personnel for foreign language training at the Army Language School. Since the construction of DLAT, the Army Language School has been incorporated into the Defense Language Institute (DLI) which offers training in over 50 different languages and now accepts both male and female students from all the military services and several other branches of government as well. Due to these program modifications and associated changes in instructional objectives, the effectiveness of DLAT as a predictor of success at DLI has come increasingly under question in recent years and has motivated governmental research into foreign language aptitude measurement.

One of the major considerations in designing a foreign language aptitude test is the type of curriculum the student will encounter (Carroll, 1962, pp. 91-94). At DLI, training is intensive and audiolingually oriented.

Students spend six hours per day in classes and have homework assignments. Classes meet five days per week for up to forty-seven week, depending upon the language, and in most DLI courses training follows Carroll (1963, pp. 1062-64) description of the "Army Method:"

1. The spoken form is presented and learned before the written form (the time lag is small).

2. The method makes use of contrastive analysis between the learner's native language and the foreign language.

3. Overlearning through "pattern practice" is stressed.

4. The desirability of simulating "real life" communication situations is well recognized.

Carroll (1962) reviewed the literature on the prediction of success in intensive foreign language training and concluded that the need for specialized tests of language aptitude arose as foreign language curriculum progressed from placing primary emphasis on translation to concentrating on a broader range of skills. Correspondingly, the effectiveness of general intelligence tests as predictors of relative success in foreign language training has been reduced as translation has lost its prime position in the foreign language curriculum.

The Harvard Language Aptitude Project, from which the Modern Language Aptitude Test (Carroll and Sapon, 1958) was developed, utilized a variety of experimental tests as predictors of success in learning a spoken foreign language (Lado, 1961, p. 374). As part of the Harvard project, Carroll (1958) conducted a factor analysis and identified seven components of foreign language aptitude which were interpreted as: Verbal Knowledge, Linguistic Interest, Associative Memory, Sound-Symbol Association, Inductive Language Ability, Grammatical Sensitivity or Syntactical Fluency, and Speed of Association.

The components of foreign language aptitude identified through earlier research provided the foundation for the present study which was designed with the following goals:

1. To produce a new test having equal or better predictive validity as compared with available commercial foreign language aptitude tests as well as with DLAT. This test would then become the successor to DLAT as a selection device for DLI. (DLI cannot consider using a commercial test as a general selection device because of lack of military control).

2. Earlier studies conducted at DLI (Asher, 1971; Al-Haik, 1972) concluded that differential prediction of success by language or language family might be achieved; so it was determined that the new test should be designed to examine the possibility of differential prediction.

3. To examine other predictor variables which might add incremental predictive validity or clarify correlational relationships (e.g. age, IQ, etc.).

Method

Predictor Variables

The following predictor variables were used:

1. Age

2. Education--scaled 1 through 9, corresponding to nonhigh school graduate through doctorate.

3. Defense Language Aptitude Test (DLAT)--a 59 item multiple-choice, paper-and-pencil test requiring the student to make translations from English into an artificial language and vice versa.

Modern Language Aptitude Test--Form A (MLAT) (Carroll and Sapon, 1958):

4. MLAT Part 1: Number Learning--43 items, auditory stimuli, requires learning numbers in a new language.

5. MLAT Part 2: Phonetic Script--30 items, auditory stimuli, requires sound-symbol association.

6. MLAT Part 3: Spelling Clues--50 items, printed stimuli, measures vocabulary and sound-symbol association.

7. MLAT Part 4: Words in Sentences--45 items, printed stimuli, measures sensitivity to grammatical structure.

8. MLAT Part 5: Paired Associates--24 items, printed stimuli, requires rapid rote learning.

9. MLAT Auditory (Total Parts 1 and 2).

10. MLAT Paper and Pencil (Total Parts 3, 4 and 5).

11. MLAT Total.

Pimsleur Language Aptitude Battery (PLAB) (Pimsleur, 1966):

12. PLAB Part 1: Past Grades (Self Report).

13. PLAB Part 2: Interest--on a five-point scale, students evaluate their interest in studying foreign languages.

14. PLAB Part 3: Vocabulary--24 items, printed stimuli, selection of English synonyms.

15. PLAB Part 4: Language Analysis--15 items, printed stimuli, English translations into an artificial language.

16. PLAB Part 5: Sound Discrimination--30 items, taped, requires ability to learn phonemic distinctions.

17. PLAB Part 6: Sound-Symbol Association--24 items, taped, requires ability to associate English sounds with their written symbols.

18. PLAB Auditory (Total Parts 5 and 6).

19. PLAB Paper and Pencil (Total Parts 3 and 4).

20. PLAB Linguistic (Total Variables 18 and 19).

21. PLAB Total.

22. Otis-Lennon IQ (Otis and Lennon, 1967).

23. Need for Social Approval (NSA) (As measured by the Marlow-Crowne (1964) Social Desirability Scale)--33 T-F items such as "I never hesitate to go out of my way to help someone."

24. Taylor Manifest Anxiety scale (TMAS) (1953)--50 T-F items such as "I do not tire quickly."

Defense Language Aptitude Battery (DLAB) (See the section on the construction of DLAB below):

25. DLAB Factor I.

26. DLAB Factor II.

27. DLAB Factor III.

28. DLAT Total.

Criterion Variable

The criterion variable was the Final Course Grade, which is based upon testing and teacher assessments made throughout the course. The criterion measure was standardized within each language before computing predictor-criterion correlations for the total sample in order to correct for possible grading biases across language departments.

The Student Sample

Data were collected on over 1,300 male DLI students before they entered training in years 1969 through 1971. Due to attrition and because of missing data, the number of students was reduced to exactly 1,000 for conducting the analysis. Of this number, 121 were dismissed due to lack of aptitude (LOA) before completing the course, leaving 879 from which validity coefficients were derived. The sample is further described in Tables 1 and 2.

Insert Table 1 about here

Insert Table 2 about here

The Construction of DLAB

Linguists at DLI, utilizing knowledge of previous research in foreign language aptitude measurement and their experience with the requirements of the DLI curriculum, constructed two experimental aptitude batteries known as Horne's Assessment of Basic Linguistic Abilities (HABLA)[2] and the Al-Haik Foreign Language Auditory Aptitude Test (AFLAAT) (Al-Haik, 1972). The items in these tests formed the pool from which DLAB was constructed. The subtests can be described as follows:

HABLA

This test requires the examinee to form language concepts from pictures. For example, at the top of a page the examinee sees a set of four pictures, each captioned with a non-English (artificial language) text. Immediately below the four pictures is a set of three additional pictures. Opposite each of these are four alternative artificial language texts. The examinee's task is to identify the text which correctly conveys what he sees in the picture by generalizing from the information gained from the top of the page.

AFLAAT

Part I--Foreign Language Sounds. This part consists of three sections:

Section 1: Utterance Identification: The examinee hears a lead utterance followed by four alternatives (similar utterances) and is required to identify the option which is identical to the lead.

Section 2: Recognition of Vowel Patterns: The examinee hears four utterances, only one of which is different in vowel pattern and is required to identify the utterance with the different vowel pattern.

Section 3: Recognition of Stress Patterns: The examinee hears four utterances which have the same number of syllables and is required to identify the one having a different stress pattern.

Part II--Foreign Language Writing or Sound-Symbol Association. The examinee hears an utterance and reads four options (in Roman Alphabet) in his test booklet. His task is to identify the option which correctly symbolizes the lead in accordance with a specified set of new rules. The primary purpose of this subtest is to measure the ability to overcome proactive inhibition utilizing old orthography.

Part III--Foreign Language Grammar. In this part of the test the examinee reads a phrase or a sentence printed in the test booklet in English. He then hears four alternative translations using the same English words with slight modifications in pronunciation and word order to accommodate new grammatical rules. His task is to identify the alternative which correctly conveys the meaning of the printed utterance according to the new rules. The new rules, which are both morphological and syntactical, pose cumulative tasks of graded difficulty (78 items). Part III contains four sections. Section 1 deals with noun and adjective agreement. Section 2 concerns the possessive form. Section 3 deals with sentence structure. And finally, Section 4 requires the use of a combination of rules introduced in the previous three sections.

Test Design

DLAB had to be constructed in accordance with a number of practical as well as theoretical considerations. To meet the requirements of the military testing system, it must be a test that is easy to administer, score and interpret, and also achieve maximum predictive validity with the smallest possible number of items. It must be

a test with maximum utility as a general selection device for all languages, yet also be capable of shedding light on the question of differential prediction. Additionally, in making the decision regarding the data analysis procedures, it was felt that, although there was no well-developed theory in the selection of experimental measures, construct validity should be a consideration as well as the establishment of predictive validity. The procedures decided upon to best meet the above goals and constraints were item factor analysis and multiple regression, the rationale being to identify maximally uncorrelated skills associated with reliable item composites. These composites, as represented by factor scales, were then correlated (using stepwise multiple regression) with the criterion, both for the entire sample and within several of the languages. For purposes of cross-validation, the total student sample was randomly divided in half (having equal numbers from each language) and the regression weights derived from one half used to predict the criterion scores of the other half and vice versa. The predicted and actual criterion scores were then correlated, resulting in two validity coefficients. The same procedure was followed using MLAT and PLAB subtests as predictors in order to obtain comparative data. These

coefficients were further compared with the zero-order correlations for the total tests on the entire sample as well as with the zero-order correlations obtained from the other predictors.

The Factor Analysis

In conducting the factor analysis, several problems were encountered, the first being the size of the item correlation matrix. There were 167 items, resulting in a 167 by 167 correlation matrix, and the storage required for factoring such a matrix was well beyond the capacity of any computer to which there was ready access. The decision was therefore made to halve the matrix by splitting on every other item (so that each half would contain the same number of items from each subtest) and factor each half separately. Factors were then to be matched on a logical basis.

The factor analysis was conducted utilizing a computer procedure developed by Guertin & Bailey (1971) from which a principal axis solution (utilizing multiple R^2 as the communality estimate) with varimax rotations is calculated from tetrachoric correlations. A criterion of at least 1.31 for latent roots (a feature of the computer program determined by the number of variables)

was set for including a factor in rotation. However, this procedure produced different numbers of factors for each matrix, and these were difficult to match logically (by examining how items "fell out" of subtests).

This difficulty was discussed with Harry Harman of Educational Testing Service (ETS) who suggested that the problem might arise from the fact that a matrix of tetrachoric correlations may not be consistent, especially since extreme dichotomies in item values affect the stability of the tetrachoric coefficient (see McNemar, 1965, p. 197). It was also learned that ETS had developed a computer procedure for matching factors from separate matrices (derived from the same population) of the type under investigation, and the decision was made to utilize this procedure. The item correlations were then recomputed using phi (although there are several problems concerning the choice of any correlational method for item factor analysis--see Lord & Novick, 1968, pp. 348-349, 537), and rearranged into a supermatrix composed of four submatrices: the original odd (84 x 84) and even (83 x 83) matrices on the diagonal, and the two covariance (item cross-correlation) matrices (84 x 83 and 83 x 84) as the off-diagonal elements.

The procedure for analyzing this supermatrix begins by subjecting each of the diagonal submatrices to a principal component analysis. Six components were retained in each set (following an arbitrary level of .50 as a minimum reliability for retaining a component). Then the correlations between these components were obtained and subjected to canonical correlation analysis across the two sets of 6 components (for a brief description of the "canonical correlation model" see Cooley and Lohnes, 1971, pp. 168-176). It was decided that only three of the canonical correlations were sufficiently large (in excess of .5), and from their associated canonical variates it was possible to determine 3 interbattery factors. From these three factors, items were selected and factor scores (they are not true factor scores in a strict sense as the next section will show) were then correlated with the criterion in the manner previously described.

Item Selection

In selecting items to be scored on each factor scale arbitrary decision rules were set. First, the item had to have a loading on the factor of .20 or higher, and second, if there was more than one loading

of that magnitude for the same item, it was grouped only by its highest loading. By this method, most items "fell out" nicely by subtest, although there were some exceptions. Only two items from AFLAAT Part I, Sections 1 and 2 reached the required factor loading and both sections were therefore eliminated from further analysis. Part II of AFLAAT was also eliminated on similar basis. All of the items from HABLA loaded on a single factor but the final 30 items in the test require the reading of new instructions for every six items. These items were therefore eliminated for the sake of brevity (as that factor accounted for enough additional items to produce sufficiently reliable subscale scores without them). Of the remaining subtests, Factor I is made up of HABLA and AFLAAT Part III, Section 4, Foreign Language Grammar--rules in combination. Factor II is composed of AFLAAT Part I, Section 3, Recognition of Stress Patterns, and Part III, Section 1, Foreign Language Grammar--Noun and Adjective Agreement. Factor III contains AFLAAT Part III, Section 2, Foreign Language Grammar--Possessive Form. One subtest, AFLAAT Part III, Section 3, Foreign Language Grammar--sentence structure, is equally distributed between the three factors. A few items were eliminated from each

of the subtests for not meeting the standard set for magnitude of factor loadings. The procedure devised for obtaining scores on each of the factor scales was to treat them as subtests, giving items a weight of 0 or 1, and rescoring student records on that basis.

Results

Insert Table 3 about here

Insert Table 4 about here

Insert Table 5 about here

Insert Table 6 about here

Insert Table 7 about here

Insert Table 8 about here

DISCUSSION

Perhaps the most notable result of this study was that, considering that the population is quite restricted in range of talent (see Table 2), the DLAB predictive validity coefficients are generally high. There is also evidence that there is considerable variance in language aptitude between students from different language departments (as indicated by differences in mean DLAB scores, Table 8), meaning that a student has unequal chances of relative success in different departments as determined by the level of ability in the group to which he is assigned. This would be expected to further attenuate the predictive validities of the aptitude measures computed on the entire sample. The reasons for the low predictive validity for the Mandarin Air Force Aural Comprehension Course (MAFAC) are not yet fully understood (Table 8), but MAFAC has a much different curriculum and different assessment procedures than other courses.

DLAB should be a superior general selection device for DLI compared to DLAT, and its predictive validity for the DLI population appears to be equal to or slightly superior to the validities achieved by PLAB and MLAT. Also, DLAB will be composed of only 108 items which should require

little more than an hour to administer. However, emphasis must be given to the fact that DLAB as a physical test product has not yet been produced and all statistics reported for the test are based on extractions from item statistics. Therefore, further validation studies will be conducted once DLAB is produced.

Factor analysis provided a useful technique for item selection, but there is little evidence that differentially weighting factors (or subtests in the case of MLAT and PLAB) will produce better predictive validities than using total scores as predictors. This is in line with Guion's (1965, pp. 166-169) review of the literature on differential weighting. As for differential prediction, that is an issue of considerable complexity. The Asher[3] (1971) study, which was conducted on some of the same data as the present study, used 37 predictor variables and 8 criteria in a series of factor analyses conducted in each language. Asher concluded from these analyses that "...each language may have a unique factor structure with very little overlap between languages...", and that this might eventually result in a greater precision for predicting how successful a student will be in each language taught at DLI. However, the ratio of variables to sample size was often as

large as 1 to 1 in these factor analyses; some of the variables were not independent; the criterion variables may not have had equal variance or reliability in each language, and students were not assigned to languages at random. But even if the assumption is made that these studies revealed meaningful differences in factor patterns, differential prediction does not seem likely to become a reality.

It does not appear that unique factor structure alone is a sufficient condition for fostering the belief that differential prediction can be achieved. The major considerations appear to be whether or not language specific factors can be isolated, and if such factors contribute substantial variance to criterion measures. The differing features of languages (such as degree of agglutination, tonal emphasis and morpho-phonemic complexity) may suggest that differing aptitudes are involved in their acquisition and use, but in areas of aptitude seemingly much more diverse than language ability, differential prediction has met with only limited success (see Cronbach, 1970, pp. 371-373). Also, Carroll (1973, p. 5) has concluded that "...(foreign language) aptitude is equally relevant to any foreign language that the individual might choose to study,..."

Cronbach (1970, p. 373) contends that the lack of success of differential prediction is probably less attributable to weakness in the predictors than the fact that examinations and marks in all areas of academic instruction are dependent upon a similar complex of abilities. And at DLI, students in different languages are not only assessed by similar means, they generally receive very similar methods of instruction as well. It therefore appears improbable that any language specific aptitudes which might exist would have a significant influence on the criterion measures in this study.

Table 8 indicates that differential weighting of DLAB subscales does not produce significantly better predictive validities within individual languages than those derived from total scores. The regression weights did show some large differences between language; but the sampling error of the weights is too large (because of small sample sizes) to make meaningful interpretations about the importance of particular abilities for success in different languages, and with the exception of MAFAC, predictive validities do not vary greatly across languages, indicating that many of the same skills are involved in the acquisition of any foreign language using the audio-lingual method of instruction.

Carroll (1974, p. 293) has observed that there does not appear to exist in the literature any systematic statement about how one goes about devising an aptitude battery, and there may be any number of ways of concocting items and combining them by various methods into a language aptitude battery. As has been noted, the production of items in this study was guided largely by previous research and theory in the field, but the use of item factor analysis to select and combine items may require some defense. An obvious question is, how can one be sure that the items eliminated by factor analysis might not provide incremental predictive validity even though they do not correlate highly with each other or with the items that were selected? To help resolve this question, a spot check of the relationship of these items with the criterion was made and they were found generally to have low to negative discrimination indices for the criterion. And as an additional check, the predictive validity of HABLA was compared with the predictive validity of the subset of items from HABLA which were selected for DLAB and they were found to be identical. It is also of interest to note that the subtests eliminated through the factor analysis appear to

be most concerned with sensory discrimination, whereas the items retained require a greater ability to learn and apply linguistic rules.

It is true that the selection of items by the magnitude of their correlations with the criterion could probably have produced a test of equal predictive validity from the item pool available but there are serious objections to this procedure, particularly when there is a concern with isolating the constructs involved in test performance (Nunnally, 1967, pp. 245-249). And such a procedure would be unduly clumsy for attempting differential prediction at DLI as it is conceivable that the result might have been a different test for every language.

The three DLAB factors do not, of course, correspond with the seven identified by Carroll (1958), but the variety of variables appears to have been wider in Carroll's study and the method of factor analysis was much different. However, Table 5 indicates that DLAB, MLAT, PLAB and DLAT all have a great deal of common core. The difficulty of making definitive interpretations of DLAB factors is also evidenced in Table 5 in that they can be seen to have complex relationships with other variables. Also, Table 4 indicates that there is a considerable amount of overlap

between the DLAB factor scales which further complicate interpretations. In short, the data are not sufficient to permit an adequate interpretation of factors, and at this stage there may be some danger in offering interpretations. Further studies will be conducted on the factor structure of DLAB in the next phase of research.

In conclusion, DLAB appears to be a test with considerable promise as a general selection device and it will be used for that purpose by each of the military services, but its utility for assigning students to specific language courses does not appear to be high, although further research on differential language aptitude will be conducted. Also, changes in selection procedures may have changed the constituency of the military population since the time the data were collected, and no data have as yet been gathered on women. Therefore, further validation research will be conducted before the test becomes operational. There are also plans to try some additional biographical and motivational items in this next phase of research, and to examine the interaction of language aptitudes with different methods of instruction as a continuing part of aptitude research at DLI.

REFERENCES

Al-Haik, A. R. Exploring the Auditory Aspects of Aptitude for Intensive Modern Foreign Language Learning. Unpublished Doctoral Dissertation, University of California, Berkeley, 1972.

Asher, J. J. Differential Prediction of Student Success in Intensive Language Training. Unpublished Governmental research report, Defense Language Institute, Presidio of Monterey, California, 1971.

Carroll, J. B. "A Factor Analysis of two Foreign Language Aptitude Batteries." Journal of General Psychology, 1958, 59, 3-9.

Carroll, J. B. "Prediction of Success in Intensive Foreign Language Training." Training Research and Education. Edited by Robert Glaser. Pittsburgh: University of Pittsburgh Press, 1962.

Carroll, J. B. "Research on Teaching Foreign Languages" Handbook of Research on Teaching. Edited by N. L. Gage. Chicago: Rand McNally, 1963.

Carroll, J. B. "Implications of Aptitude Test Research and Psycholinguistic Theory for Foreign Language Teaching." International Journal of Psycholinguistics, 1973, 2, 5-15.

Carroll, J. B. "The Aptitude-Achievement Distinction: The Case of Foreign Language Aptitude and Proficiency." The Aptitude-Achievement Distinction. Edited by D. R. Green. Monterey, California: CTB/McGraw-Hill, 1974.

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

Cooley, W. W., & Lohnes, P. R. Multivariate Data Analysis. New York: Wiley & Sons, 1971.

Cronbach, L. J. Essentials of Psychological Testing (3rd ed.) New York: Harper & Row, 1970.

Crown, D. P., & Marlowe, D. The Approval Motive: Studies in Evaluative Dependence. New York: Wiley & Sons, 1964.

Guertin, W. H., & Bailey, J. P. Introduction to Modern Factor Analysis. Ann Arbor: Edwards Bros., 1970.

Guion, R. M. Personnel Testing. New York: McGraw-Hill, 1965.

Horne, K. M. "Differential Prediction of Foreign Language Testing." Paper presented before the Bureau of International Language Coordination, London, June, 1971.

Lado, R. Language Testing. New York: McGraw-Hill, 1968.

Lord, F. M., & Novick, M. R. Statistical Theories of Mental Test Scores. Reading, Massachusetts: Addison-Wesley, 1968.

McNemar, Q. Psychological Statistics. (3rd ed.) New York: Wiley & Sons, 1965.

Nunnally, J. C. Psychometric Theory. New York: McGraw-Hill, 1967.

Otis, A. S., & Lennon, R. T. Otis-Lennon Mental Ability Test, Advanced Level, Form J. New York: Harcourt, Brace, and World, 1967.

Pimsleur, P. Pimsleur Language Aptitude Battery, Form S. New York: Harcourt, Brace, and World, 1966.

Taylor, Janet A. "A Personality Scale of Manifest Anxiety." The Journal of Abnormal and Social Psychology XLVIII, 1953, 2, 285-290.

Footnotes

[1]Appreciation is expressed to Harry H. Harman, Director, Developmental Research Division, Educational Testing Service, and John B. Carroll, Director, L.L. Thurstone Psychometric Laboratory, University of North Carolina, for their suggestions in the preparation of this research report.

[2]Developed by Colonel (Dr.) Kibbey M. Horne, a linguist and former Commandant of the Defense Language Institute, West Coast Branch from 1968-1972, who was instrumental in beginning aptitude research at DLI (see Horne, 1971). He is presently Director, International Programs, California State University and Colleges.

[3]Dr. James Asher, Professor of Psychology, San Jose State University, who conducted this research study under contract to the Defense Language Institute, West Coast Branch.

TABLE 1

Sample Description

Language	Length of Course (Weeks)	Number of Students LOA	Graduated	TOTAL
Arabic	47	10	112	122
Chinese-Mandarin	47	8	70	78
German	32	1	87	88
Korean	47	11	54	65
Russian Aural Comprehension	37	23	75	98
Russian Basic	47	12	111	123
Spanish	24	3	87	90
Turkish	47	6	40	46
Vietnamese	47	8	59	67
French	24	9	63	72
Chinese-Mandarin Air Force Aural Comprehension (MAFAC)	32	30	121	151
TOTALS		121	879	1,000

TABLE 2

Sample Norms

	LOA	Graduated	TOTAL
Mean Age	21.6	21.9	21.86
Education (mean years of college attendance)	1.5	2.6	2.4
Otis-Lennon IQ (12th Grade Norms)	114	120	119
Mean MLAT (percentile freshman male norms)	49	75	70
Mean PLAB (percentile end first course high school norms)	82	95	95

TABLE 3

DLAB Statistics

Subtest	Number of Items	Mean Score	Std Dev	Reliability KR-21
Factor I	49	29.76	7.57	.81
Factor II	36	26.07	5.48	.78
Factor III	23	18.74	4.02	.82
	108			

NOTE.--KR-21 generally underestimates KR-20 but was computed instead to avoid complex programming requirements.

TABLE 4

Intercorrelations of Factor Scales

	I	II	III
Factor I	1.0		
Factor II	.374	1.0	
Factor III	.540	.466	1.0

TABLE 5

Highest Correlations of DLAB Subscales

With Other Predictors

Variable	Factor I
11 MLAT-Total	.575
7 MLAT-Part 4: Words in Sentences	.572
10 MLAT-Paper and Pencil	.555
22 Otis-Lennon IQ	.555
15 PLAB-Part 4: Language Analysis	.544
3 DLAT	.496
21 PLAB-Total	.495
Variable	**Factor II**
20 PLAB-Linguistic	.570
17 PLAB-Part 6: Sound Symbol Assoc.	.560
5 MLAT-Part 2: Phonetic Script	.538
21 PLAB-Total	.528
11 MLAT-Total	.490
16 PLAB-Part 5: Sound Discrimination	.478

(Table 5 con't)

Variable	Factor III
11 MLAT-Total	.476
20 PLAB-Linguistic	.470
22 Otis-Lennon IQ	.456
5 MLAT-Part 2: Phonetic Script	.441
7 MLAT-Part 4: Words in Sentences	.441

TABLE 6

Correlations of Predictors With Average Grade (Standardized Within Language) For "Graduated" Sample (879 Cases):

Predictor	Correlation
1. Age	.029
2. Education	.185
3. DLAT	.373
4. MLAT-Part 1: Number Learning	.155
5. MLAT-Part 2: Phonetic Script	.307
6. MLAT-Part 3: Spelling Clues	.324
7. MLAT-Part 4: Words in Sentences	.359
8. MLAT-Part 5: Paired Associates	.243
9. MLAT-Auditory (Total 4 & 5)	.266
10. MLAT-Paper and Pencil (Total 6, 7 & 8)	.413
11. MLAT-Total	.401
12. PLAB-Part 1: Past Grades (Biographical)	.258
13. PLAB-Part 2: Interest	.113
14. PLAB-Part 3: Vocabulary	.267
15. PLAB-Part 4: Language Analysis	.263
16. PLAB-Part 5: Sound Discrimination	.198
17. PLAB-Part 6: Sound Symbol Assoc.	.204

(Table 6 con't)

Predictor	Correlation
18. PLAB-Auditory (Total 16 & 17)	.253
19. PLAB-Paper and Pencil (Total 14 & 15)	.319
20. PLAB-Linguistic (Total 18 & 19)	.357
21. PLAB-Total	.405
22. Otis-Lennon (IQ)	.272
23. Need for Social Approval Scale	-.078
24. Taylor Manifest Anxiety Scale	-.040
25. DLAB-Factor I	.384
26. DLAB-Factor II	.317
27. DLAB-Factor III	.315
28. DLAB-Total	.431

TABLE 7

Cross-Validations

	Multiple Correlation (uncorrected for shrinkage)		Correlation of Predicted and Actual Criterion Scores	
	Sample		Sample	
	1	2	1	2
MLAT (5 subtests as predictors)	.464	.444	.436	.420
PLAB (6 subtests as predictors)	.473	.405	.415	.347
DLAB (3 subscales as predictors)	.448	.423	.441	.413

NOTE.--Weights derived from the multiple regressions on each half of the sample are used to predict the criterion scores on the other half in a double cross-validation design. Thus, for example, the multiple correlation for MLAT obtained in Sample 1 (.464) is to be compared with the correlation of

(Table 7, Note con't)

predicted and actual criterion scores (.420) obtained in Sample 2. Similarly, the multiple correlation (.444) obtained in Sample 2 is to be compared with the correlation (.436) obtained in Sample 1.

TABLE 8

DLAB Predictive Validities Within Six Languages

Language	N	DLAB Total Mean Score	Multiple Correlation or DLAB Subscales with Criterion	Zero Order Correlation of DLAB Total Scores with Criterion
Arabic	112	75.4	.564	.562
Chinese-Mandarin	70	75.9	.529	.485
German	87	74.2	.548	.543
Russian Basic	111	77.9	.501	.497
Spanish	87	71.3	.570	.554
MAFAC	121	77.0	.198	.194

DEFENSE LANGUAGE INSTITUTE
OFFICE OF THE DIRECTOR
PRESIDIO OF MONTEREY, CALIFORNIA 93940

REPLY TO
ATTENTION OF: DLIF-D

30 MAY 1975

SUBJECT: Proposed Change in Defense Language Proficiency Test Scoring

HQDA (DAPE-MT)
Washington, D.C. 20310

1. References:

a. AR 350-20, OPNAVINST 1550.7, AFR 50-40, MCO 1550.4B, 17 Jan 74, Management of the Defense Language Program.

b. DAMI-OC, MEMORANDUM FOR: DEPUTY CHIEF OF STAFF FOR PERSONNEL, 13 Mar 75, Subject: Army Linguists.

c. DF, DAPE-MPT, Subject: Arabic Linguists (title unclassified), 18 Apr 75.

d. See Inclosure 1 for additional references.

2. DLI is considering a change in the method of scoring Defense Language Proficiency Tests (DLPT) which will provide more valid and precise information regarding the foreign-language ability of personnel assigned to, or to be selected for, positions requiring language ability. This change may have far-reaching implications involving several major staff agencies in each Service, including intelligence, recruiting, training, and several aspects of military personnel management. The purpose of this letter is to request reactions regarding the possible impact of this proposal on each of the Services.

3. Measurement of foreign-language ability is an essential aspect of management of DLI and of the entire Defense Language Program, since determinations of capability and potential utilization necessarily rely heavily on a testing system. This proposal is therefore related to such current

actions as The Army Linguist Personnel Study (TALPS), (ref 1b), the Arabic linguist study (ref 1c), and the current GAO survey of Governmental language training, in which one major concern is with language qualifications of personnel assigned overseas.

4. While the plan described below incorporates certain technical improvements, the change that would be most apparent to the Services will be in the way DLPT scores are derived and reported. The present procedure is to convert raw scores to skill categories, called "levels," ranging from level-5 (proficiency equal to that of a native speaker) down to level-0 (no proficiency), by means of a conversion chart. Thus, for example, all linguists whose raw scores on the listening comprehension part of the DLPT I are in the range 40 to 47 are now simply categorized as level-3, in listening ability. Since few trainees achieve higher than level-3, and few are as low as level-1 at graduation, the present system does little more than divide them into two groups: level-2 and level-3. Under the proposed procedure, raw scores would be converted to a standard scale (T-scale). Listening comprehension skills of individual linguists would, under this plan, be recorded as T-scores, such as 39, 41, 47, 71, etc., permitting far more precise identification of the relative degree of foreign-language comprehension possessed by each.

5. The reasons for proposing a change in DLPT scoring are both technical and managerial. The technical reasons make it impossible for DLI to assure the accuracy of conversions from DLPT or DLPT II scores to skill levels. (The technical psychometric problems are briefly summarized in Inclosure 2, which may be of interest to your research specialists.) For the next two or three years, we will continue to publish tables for conversion of DLPT scores to skill levels, but with a cautionary notation that they are not statistically validated.

6. The managerial reasons derive from the technical problems, from budgetary limitations, and from our desire to provide you with the best available data upon which to base your decisions about job assignments, promotions, retention, etc., where ability in a foreign language is a factor. The proposed system will provide you with the following:

a. Instead of gross categorization of skill level (1, 2, or 3), it will provide a continuous scale which will permit the ranking of personnel by language ability with far greater accuracy.

b. Instead of an unvalidated conversion to levels, it will provide for a statistically determined conversion, i.e., to a standard scale.

c. Instead of oral-proficiency ratings (S-ratings) derived with questionable accuracy from DLPT scores, it will provide for such ratings when, but only when, they have been determined by means of the oral interview technique.

d. Instead of a set of tests that must remain in use over many years thereby suffering from compromise and obsolescence, it is anticipated that the new system will permit replacement of compromised and/or obsolete tests on a much more timely basis.

7. In the interim, new DLPTs (designated DLPT II) are being introduced into the system through HQDA (DAPC-PME-T), AUTOVON 221-9750. Printing and distribution is done by TAGO. The first eight of these DLPT IIs were introduced in 1974. The next fifteen DLPT IIs, plus four reading proficiency tests (DLRPTs), are currently being reproduced by the Adjutant General's office and will be introduced later this year. The scoring keys on these nineteen tests have raw-score to T-score conversion tables and are the first that contain the cautionary statement that conversion to skill levels has not been statistically validated.

8. The new procedure would be implemented in stages.

a. Provision of T-score conversions in addition to skill-level conversions has already begun (para 7, above).

b. In approximately two years, the first few tests of the new type would be introduced with T-score conversions only.

c. At that time, we propose to discontinue conversion to skill levels from all other tests then in the system for which T-scale conversion data are available.

d. The remaining tests would be implemented, several at a time, to replace the old ones, as they are developed.

e. It is our present intention to introduce all new tests through the same procedures we have used in the past, that is, through DAPC-PME-T and TAGO as noted in paragraph 7 above.

(47)

DLIR-D 30 MAY 1975
SUBJECT: Proposed Change in Defense Language Proficiency Test Scoring

9. The new method would require that your regulations, personnel records, and so on, be revised to employ the two-digit standard scores for DLPT scores, instead of the present notations of skill levels. (No one-digit or three-digit standard scores will ever occur.) Provision for notations of oral-proficiency levels (S-2, S-3+, etc.) should be retained, however, for those cases in which S-ratings are determined by interview. These changes would be required at the time we would discontinue publishing level-conversions for DLPTs, in about two years.

10. If this new method is adopted, DLI will provide recommendations regarding changes of wording to relevant documents in each Service. An alternative, on which we would like your comments, would be publication of one or more joint regulations covering some of these matters (as is the case for Reference 1a).

11. For your convenience, we have listed (Inclosure 1) all the manuals, regulations, etc., known to us that relate to the testing or management of Army, Navy, Air Force, and Marine Corps language-qualified personnel. We would appreciate your adding to this list, in your response, any others we have overlooked.

12. The proposed plan will not alter in any way our separate research and development efforts on criterion-referenced tests for DLI students from the Security Services.

13. Your response is requested within 60 days.

14. (Point of contact: Dr. Francis A. Cartier, AUTOVON 973-8443.)

2 Incl
as

JAMES R. KOENIG
Colonel, USA
Director

CF:
HQ TRADOC, ATTN: ATTNG (LTC Weston)
DAPC-PME-T

(48)

Technical Considerations in Changing Defense Language Proficiency Test Scoring

The present system for scoring and interpretation of DLPTs was established more than twenty years ago and reflects the state of the art in language testing at that time. The validity of this system has come increasingly under question in recent years, causing DLI to examine a number of alternatives.

The basic problem concerns the use of DLPT scores to derive ("predict") language proficiency "levels" (defined by narrative descriptions of performance in real-life situations). However, quantitative criterion data representative of the level descriptions has never been collected. Hence, there is no psychometrically acceptable evidence that persons who achieve a particular DLPT score will, in fact, be capable of performing at a particular performance level in real-life circumstances. It is apparent, then, that the present method of DLPT score interpretation is inconsistent with generally-accepted validation procedures (e.g., as set forth in the APA/AERA/NCME Standards for Educational and Psychological Tests.)

The possibility of validating DLPTs against real-life criteria has been examined, but numerous practical and theoretical problems, as well as budgetary and manpower limitations, prohibit this alternative. For example, the small number of students trained annually in most languages at DLI precludes all but a rudimentary examination of criterion-related validity in most languages for which DLPTs currently exist. Additionally, the level-of-performance descriptions possess inherent limitations as the basis for a criterion metric (e.g., descriptions such as "minimum professional competency" are not easily quantified).

Nevertheless, the DLPT does provide reliable ordinal information. DLI is therefore proposing a method of DLPT scoring and interpretation which is consistent with that information. The method appropriate to the available data is the standard score; we therefore propose to employ the T-scale ($\bar{x}$ = 50, S.D. = 10). See the example conversion table on the next page.

One rationale for this proposal is that standard scores are widely understood and used within the Armed Services. It should therefore be an easy matter for the Services to tie such actions as job assignments, promotions, pay

SAMPLE DLPT II FORM A CONVERSION TABLE

LISTENING COMPREHENSION

RS-CS	RS-CS	RS-CS	RS-CS
0=00	16=26	32=46	48=66
1=00	17=28	33=47	49=68
2=00	18=29	34=49	50=69
3=00	19=30	35=50	51=70
4=00	20=31	36=51	52=71
5=00	21=32	37=52	53=73
6=00	22=34	38=54	54=74
7=00	23=35	39=55	55=75
8=00	24=36	40=56	56=76
9=00	25=38	41=57	57=78
10=00	26=39	42=59	58=79
11=00	27=40	43=60	59=80
12=00	28=41	44=61	60=81
13=00	29=42	45=63	
14=00	30=44	46=64	
15=00	31=45	47=65	SEM = 3.5

READING COMPREHENSION

RS-CS	RS-CS	RS-CS	RS-CS
0=00	16=26	32=46	48=66
1=00	17=28	33=47	49=68
2=00	18=29	34=49	50=69
3=00	19=30	35=50	51=70
4=00	20=31	36=51	52=71
5=00	21=32	37=52	53=73
6=00	22=34	38=54	54=74
7=00	23=35	39=55	55=75
8=00	24=36	40=56	56=76
9=00	25=38	41=57	57=78
10=00	26=39	42=59	58=79
11=00	27=40	43=60	59=80
12=00	28=41	44=61	60=81
13=00	29=42	45=63	
14=00	30=44	46=64	
15=00	31=45	47=65	SEM = 3.5

LEGEND: RS = RAW SCORE, CS = CONVERTED SCORE, SEM = STANDARD ERROR OF MEASUREMENT (IN RAW SCORE POINTS)

NOTES:

1. The converted score is T-score.
2. Conversion of T-scores into proficiency levels is not yet statistically validated.

Example of chart for conversion of DLPT II raw scores (RS) to converted scores (CS).

increases, etc., to specific standard scores instead of to proficiency levels. Furthermore, standard scores will provide a means for making much finer discriminations among personnel than do proficiency levels (which, in effect, provide only a 3-point scale for differentiating among DLI graduates, and, at best, a 5-point scale).

When DLI began introducing a new series of DLPTs (DLPT II) in 1974, the first eight tests proved to have considerably lower means than the tests they replaced. Lacking true criterion data, and faced with the necessity of equating the two forms, DLI could only provide a raw-score conversion system which allows comparable proportions of students to achieve each "level." The validity of this procedure is, of course, not acceptable, especially since the old tests were suspected of being compromised, but it was regarded as a short-term solution pending implementation of the system of standard scores mentioned above. With the exception of the first DLPT IIs introduced (Arabic-Egyptian, Arabic-Syrian, Chinese-Mandarin, Russian, Spanish-Latin American, Thai, Vietnamese-Hanoi, and Vietnamese-Saigon) all remaining DLPT IIs (currently being reproduced by the Adjutant General's Office) will have converted scores which are, in fact, T-scores. The latter tests will also contain a provision for converting the T-scores to "levels," but a statement on the scoring key informs the user that the conversion to proficiency levels has not been statistically validated.

DLI also has plans for revising the method of developing future DLPTs. The emphasis will change from attempting to measure language proficiency as a general trait to developing tests which are content valid with respect to DLI courses. Among the advantages of the content validation model are: test content will be more closely tied to Defense Language Program training objectives; it eliminates the necessity for collecting the extensive statistical data necessary for criterion-referenced validation (data which in any case, is not presently available to DLI); alternate forms can be more easily equated on a linguistic-content basis when the number of examinees is too small for producing statistically parallel tests; DLI instructors can be better utilized as item writers because of their knowledge of course content; and test content is tied to a definite content domain instead of the abstract, and presently ill-defined, domain of "general language proficiency". These new tests would also be scored on a standard score basis.

MILITARY DIRECTIVES DIRECTLY AFFECTED BY FOREIGN LANGUAGE PROFICIENCY TESTING

DLI maintains current files only on Army publications. The Air Force, Navy and Marine Corps publications may not be accurately listed. Corrections and additions will be welcomed.

Military Service	Publication	Basic Date	Title
Army	AR 350-20	17 Jan 74	Management of the Defense
Air Force	AFR 50-40	17 Jan 74	Language Program
Navy	OPNAVINST 1550.7	17 Jan 74	
Marine Corps	MCO 1550.4B	17 Jan 74	
Army	AR 611-6	3 Nov 69	Army Linguist Program
Army	DA Pam 310-8	9 Jan 75	Index of Army Personnel Tests and Measures
Army	DA Pam 611-300	20 Jun 69	Administering and Scoring the Defense Language Proficiency Tests
Army	DA Cir 611-27	10 Jan 74	Introduction of Defense Language Proficiency Tests II (DLPT II)
Army	DLI Pam 611-1	1 Jun 73	The Defense Language Proficiency Tests II Administration and Scoring Manual
Army	DLI Pam 611-2	1 Feb 75 (in press)	The Defense Language Reading Proficiency Tests Administration and Scoring Manual

ilitary Service	Publication	Basic Date	Title
ir Force	AFR 0-7	16 Dec 74	Index of Air Force Personnel Tests
ir Force	AFM 35-8	24 Sep 71	Air Force Military Personnel Testing System
ir Force	AFPT 800	1 Jan 63	Administering and Scoring Language Proficiency Tests
avy	OPNAVINST 1520.17A	21 Jul 72	Foreign Language Instruction
avy	BUPERINST 1550.42	10 Feb 65	Foreign-Language Aptitude and Proficiency Tests
SMC	MCO P1000.6B	10 Dec 73	ACTS Manual

MILITARY DIRECTIVES INDIRECTLY AFFECTED
BY FOREIGN LANGUAGE PROFICIENCY TESTING

ilitary Service	Publication	Basic Date	Title
epartment of Defense	Directive No. 5160.41	5 Oct 68	Defense Language Program
rmy	AR 340-16	26 Jun 73	Safeguarding "FOR OFFICIAL USE ONLY" Information
rmy	AR 600-200	24 Mar 65	Enlisted Personnel Management System
rmy	AR 611-5	10 Mar 69	Army Personnel Tests
rmy	AR 612-201	30 Jun 72	Processing Procedures at U.S. Army Reception Stations and Training Centers and Control and Distribution of Trainees
rmy	AR 621-5	5 Nov 64	General Educational Development
ir Force	AFR 205-1	1 Feb 73	Information Security Program
ir Force	AFR 205-13	10 Oct 73	Safeguarding Controlled Item (Test Material) Information
ir Force	AFM 12-50	20 Dec 72	Disposition of Air Force Documentation
ir Force	AFM 35-14	1 Feb 71	Military Personnel Records System
avy	BUPERSINST 1550.41A	11 Jun 69	Unknown

IN REPLY REFER TO
Ser 991E5/71613

10 JUN 1975

From: Chief of Naval Operations
To: Deputy Chief of Staff for Personnel, Department of the Army (DAPE-MPT)

Subj: Management of the Defense Language Program; request for change to

Ref: (a) Army Regulation No. 350-20 (OPNAVINST 1550.7)
(b) Manual of Navy Officer Classifications NAVPERS 15839B (NOTAL)

1. Reference (a) provides the policy and procedures for the management of the Defense Language Program. It also lists the four functional skill areas of language as Listening Comprehension (C), Reading (R), Speaking (S) and Writing (W). Reference (b) is the Navy's regulation that applies the DOD standardized language codes and definitions of proficiency levels. It also contains a prescribed data chain so that a given language and proficiency level in each of the four functional skill areas may be coded in a six character code, i.e., FR2212. The data chain requires that the four functional skill areas be listed in the order indicated above.

2. Reference (a) lists the four functional skill areas, in different sequences, in paragraphs 2-4d(4), 5-1, and 5-4. For the purpose of consistency and to facilitate interpretation of manpower and personnel documents, it is requested that the functional skill areas be listed wherever appearing in future editions of reference (a) in the following order: C, R, S, and W.

R. L. Harf
By direction

(55)

CONVERSIONS OF T-SCORES TO PERCENTILES

T-Score	Percentile	T-Score	Percentile
75 and above	99	51	53
74	99	50	50
73	98	49	46
72	98	48	42
71	98	47	38
70	97	46	34
69	97	45	30
68	96	44	27
67	95	43	24
66	94	42	21
65	93	41	18
64	91	40	15
63	90	39	13
62	88	38	11
61	86	37	9
60	84	36	8
59	81	35	6
58	78	34	5
57	75	33	4
56	72	32	3
55	69	31	2
54	65	30	2
53	61	29 and below	1
52	57		

There will have to be an evaluation made as to how these Standard Scores can be used to indicate a desired level of proficiency such as that described in the current system 1 thru 5.

(54)