

FOUR EXPERIMENTS ON THE RELATION BETWEEN FOREIGN LANGUAGE TEACHING AND LEARNING*

John A. Upshur
University of Michigan

University level students studying English as a foreign language participated in four language learning experiments which were conducted within the framework of existing EFL courses. The experiments indicated (1) that language learning is not related to amount of formal language instruction for those students concurrently enrolled in academic classes, and (2) that sequential mastery of materials is not necessary for learning in an intensive foreign language program. Even the possibility that sentences of a foreign language can be mastered in any order in which they are presented is questioned. The findings were interpreted in support of the view that the most efficient foreign language learning is informal and occurs when the learner must make communicative use of the language variety to be learned, and that the internal structure underlying a set of sentences of a foreign language is not completely learned by presentation and practice of that set of sentences.

This report seeks to provide partial answers to three related questions relevant to foreign language learning and current practices in foreign language teaching.

1. Is formal foreign language instruction useful for students living and working in that language environment?
2. Is sequential "mastery" of the materials in a foreign language course necessary?
3. Is sequential "mastery" possible?

The first question is meaningful in a number of contexts. Should foreign student advisors, for example, counsel supplementary language courses for foreign students, or can they safely assume that language learning will occur through the students' use of the language in regular course work just as well as it will in formal language courses? In practical terms the second and third questions

*Each of the reported experiments reflects, to a greater or lesser degree, the constraints imposed upon the experimenter who must work within the limitations of existing courses; there are bounds beyond which tuition paying students--and the directors and teachers of their courses--cannot be expected to stray. It would be remiss, however, not to acknowledge especially the latitude and cooperation afforded by the staff of the English Language Institute.

are especially relevant to procedures for placement in foreign language courses, to ordering of materials in course design, and to programmed language instruction. In a more theoretical frame of reference, a conclusive answer to the third question could have far reaching implications about the nature of language functioning and/or language learning.

Four experiments were conducted to provide tentative answers to these questions. In each case the foreign language of instruction was English, and the environment a university in the United States. Experiment I was designed to test the hypothesis (stated here in the null) that English language learning is not related to amount of English language instruction for students in an English language environment. Experiments II and III were designed to test the hypothesis (again stated in the null) that mastery of earlier course materials is unrelated to mastery of subsequent course materials. Experiment IV was designed to provide some data which would suggest an answer to the question of whether sequential mastery of foreign language materials is a reasonable expectation. It is clearly impossible to test the hypothesis that sequential mastery is impossible, for such a test would involve an infinite number of experiments in order to examine every possible sequence of materials in a course. And that impossible procedure would have to be carried out for courses with all possible contents.

Experiment I

For a period of seven weeks participants in the 1966 Orientation Program in American Law either received no instruction in English as a foreign language or were given instruction either one hour or two hours daily. Placement in treatment groups (0-Hr, 1-Hr, 2-Hr) was determined by an initial language proficiency test.¹ At the end of the period a parallel form of the test was administered. In addition to the language classes in which some participants were enrolled, all participants enrolled in seminars in American Law and attended a series of lectures. The 0-Hr and 1-Hr groups spent two hours daily in seminars; the 2-Hr group spent one hour daily. In the seminars for the 2-Hr group an attempt was made to compensate for the more limited language abilities of the participants. Less information was presented in each seminar hour, and the outside reading requirements were reduced.

¹J. A. Upshur, *et al.*, *Michigan Test of English Language Proficiency*. From B revised, Ann Arbor, English Language Institute, 1965.

In an earlier report² it was shown that (a) English language ability improved more for the 2-Hr group than for the 1-Hr group, and (b) that within groups seminar grades were positively related to English language test scores. In this report, however, the issue is whether improvement in language ability is attributable to amount of formal language instruction, and is not simply a function of initial proficiency.

For analysis the following sub-groups were selected: (1) the class section of ten participants in the 2-Hr group scoring highest on the initial proficiency test, (2) ten of the class section of the 1-Hr group scoring lowest on the initial test, (3) ten of the class section of the 0-Hr group scoring lowest on the initial test. The 1-Hr and 2-Hr groups were selected in this way so that initial test scores would be as similar as possible within the limitations imposed by proficiency based sectioning. Table 1 presents a summary of initial test (I) and final test (F) scores for these three sub-groups.

TABLE 1
Initial and Final English Test Scores for OPAL Participants

Group	N	Mean _I	S.D. _I	Mean _F	S.D. _F
0-Hr	10	94.00	1.18	94.20	2.60
1-Hr	10	78.10	3.94	83.10	5.84
2-Hr	10	76.10	4.78	81.30	6.72

Results

A significant difference among group means was found for both initial and final tests. (With $df = 2, 27$, $F_I = 65.26$ and $F_F = 15.30$.) Both differences are significant at the 1% level. Groups 1-Hr and 2-Hr were not significantly different, however, on either initial or final tests. (With $df = 1, 18$, $F_I = 0.97$ and $F_F = 0.38$.)

An analysis of covariance, Table 2, was performed with the three groups in order to "partial out" the effects of initial language proficiency. No significant effects on language learning attributable to amount of language instruction were found.

²J. A. Upshur, "OPAL English Language Report" submitted to the Orientation Program in American Law, September, 1966.

TABLE 2
Analysis of Covariance for OPAL Participants

Source	df	SS _I	SS _F	SI _F	SS _{F.I}	MS _{F.I}	SD _{F.I}
Among Groups	2	1924	976	1370	12	6	
Within Groups	26	398	861	357	541	20.8	4.56

$F_{F.I} = .28$ (not significant)

The adjusted means for the final English tests for the three groups were:

0-Hr = 84.09
1-Hr = 87.26
2-Hr = 87.05

In order to approach significance ($P = .10$) differences between means would have to be 3.49. Initial test scores of the 0-Hr group were so high that no measurable learning with the test used in this study was likely to occur. Therefore, the adjusted means of the 0-Hr group is somewhat artificially depressed. Yet the greatest difference between means, between groups 0-Hr and 1-Hr, still does not approach significance.

On the basis of these data it is not possible to reject the null hypothesis that language learning in the environment of that language is unrelated to the amount of formal language instruction.

Experiment II

Fifty students in the Intensive Course in English at the English Language Institute, University of Michigan, were given tests of English grammar and listening comprehension before and after a period of instruction. The instructional materials were the first ten lessons each of *English Sentence Patterns*³ and *English Pattern Practices*.⁴ The period of instruction ranged from two to three and one half weeks. The instructors teaching the different classes in which these students were enrolled determined the period for their classes. Each instructor was free to pace his instruction according to whatever criteria he might choose. The pre- and post-tests both included the *ELI English Achievement Series Gram-*

³English Language Institute Staff, Ann Arbor, The University of Michigan Press, 1957.

⁴English Language Institute Staff, Ann Arbor, The University of Michigan Press, 1957.

mar Examination A (Gr.A) and the *ELI English Achievement Series Aural Comprehension Examination A* (A-C.A).⁵ These are each fifty item achievement tests sampling from the first ten lessons of the two texts used in this experiment. The purpose of the experiment was to demonstrate that learning of previously unknown material in the first ten lessons of the instructional materials does take place during the time that the materials are being taught. The null hypothesis states that post-test scores will not differ significantly from pre-test scores. The results of Experiment II are found in Table 3. Post-test scores were higher than pre-test scores in grammar and aural comprehension. Both differences were significant at the 1% level. The null hypothesis was rejected, so from these data one can safely accept the hypothesis that some learning of the linguistic materials being taught does take place during the time they are being taught.

TABLE 3

English Language Learning Pre-test and Post-test Results

Test	N	Pre-test		Post-test		t	P
		Mean	S.D.	Mean	S.D.		
Gr.A	50	39.42	7.25	45.62	3.22	8.40	<.01
A-C.A.	50	33.37	7.76	40.82	8.87	8.19	<.01

Experiment III

Two groups, each of eleven students of English as a foreign language at the English Language Institute, were matched on the same initial tests of grammar and listening comprehension that were used in Experiment II. Three students withdrew from the Institute before the end of the experiment leaving eight matched pairs for analysis. The efficiency of matching is indicated by the correlation of initial test scores between pairs in the two groups. The coefficient of correlation (r) of initial grammar scores was .96, of initial listening test scores, .99.

Instruction for the control group began with "Lesson One" of *English Sentence Patterns* and *English Pattern Practices* and continued to the completion of the texts. Instruction for the experimental group began with "Lesson Eleven" of the texts and con-

⁵Paul W. Pillsbury, Randolph Thrasher, John Upshur, Ann Arbor, English Language Institute, 1963.

tinued to the same point. The experimental group did, however, spend two days on drills from the first ten lessons of the texts. These drills are introductions to visual aids used after "Lesson Eleven." They are designed to teach the names of objects and relations pictured on charts which are used for classroom drill. In the course of this "introduction" the experimental group had to drill three grammatical "patterns." These "patterns" were not, however, determined by errors on the initial tests.

Both groups received the same instruction in pronunciation and vocabulary classes concurrent with their different treatment in grammar and drill classes. Both groups had the same instructors, who devoted the same amount of instruction time per lesson to each group. The control group reached the end of the two texts in forty-one class days with one hour per class day given to each text. The experimental group reached the end of *English Sentence Patterns* in thirty-four days, and the end of *English Pattern Practices* in thirty-six days.

On the day following the completion of a text, subjects took a one hundred item achievement test which samples uniformly from the relevant text beginning at "Lesson Eleven." Thus the control group received the final grammar test (Gr.F)⁶ and the final listening comprehension test (A-C.F)⁷ on the forty-second class day; the control group received Gr.F on the thirty-fifth class day and A-C.F on the thirty-seventh class day.

It was shown in Experiment II above that learning of materials in "Lesson One" through "Lesson Ten" does occur while those lessons are being taught. Experiment III is concerned with performance on later materials when the initial lessons have or have not been taught. Specifically, the experiment tests the null hypothesis that a group of English language learners who have been taught earlier course materials will not perform better on a test of later course materials than a group who have not been taught these earlier materials.

Results

Table 4 presents a summary of initial and final test results for the control and experimental groups, and the results of t-tests for the significance of difference between final test means.

⁶Testing and Certification Division, *Final Examination-Grammar*, Ann Arbor, English Language Institute, n.d. (1961).

⁷Testing and Certification Division, *Final Examination-Aural Comprehension*, Ann Arbor, English Language Institute, n.d. (1961).

TABLE 4

Initial and Final Test Performance for
Total-Study and Partial Study Language Learners

Test	Control Group		Experimental Group		t	Significance Level
	Mean	S.D.	Mean	S.D.		
Gr.A.	46.25	2.64	46.13	2.32	---	---
A-C.A.	43.13	2.09	43.13	2.42	---	---
Gr.F	94.12	3.69	95.62	2.60	-.90	N.S.
A-C.F	86.50	2.86	88.38	5.70	-.58	N.S.

Final test means of the control and experimental groups were very close on both tests. The slight difference is far from significant, but the direction of the difference is opposite to the "expected" outcome. These results do not allow rejection of the null hypothesis.

Experiment IV

Experiment III at least raises some doubt as to whether sequential mastery of "items" is necessary even in a course in which new "items" are presented as contrasts to, and elaborations of, previously learned items. Experiment IV was conducted in order to gain more information about the order of mastery of items in a foreign language course.

Fifty-two English language students were enrolled in five class sections ranging in size from nine to twelve students each. Class sections were relatively homogeneous with respect to English language ability. Class sectioning was based upon initial grammar and listening comprehension tests.

After twenty-one days of intensive (five hours per day) study a one hundred item grammar test and a one hundred item listening comprehension test were administered. Test items were based upon, and ordered corresponding to, the linguistic "items" in the instructional texts, *English Sentence Patterns* and *English Pattern Practices*. At the time of testing the five class sections had all received instruction on different amounts of the instructional texts. In *English Sentence Patterns* the slowest class section had received instruction on materials covered by the first thirty-eight items of the grammar test; the fastest section had received instruction on materials covered by the first ninety-six test items. Corresponding figures for drill in *English Pattern Practices* were thirty-nine

and eighty-three items of the listening comprehension test. Although the amount of text material taught in each class was different, the order of instruction was the same. Instructors were responsible for the teaching rate in their own classes. With one exception the teaching rate preserved the order of the different initial mean abilities of the classes. The class with the second highest initial mean test scores had the most material presented.

Mean class scores on the two achievement tests are shown in Figure 1. The number of test items answered correctly is plotted as a function of the amount of text material taught. As a rule classes which have been taught more text material performed better on the achievement tests than classes which have been taught less. ($Rho = .90$ for grammar, $Rho = 1.00$ for listening; $P < .05$ with $df = 3$.) It is not possible to infer a causal relationship here,

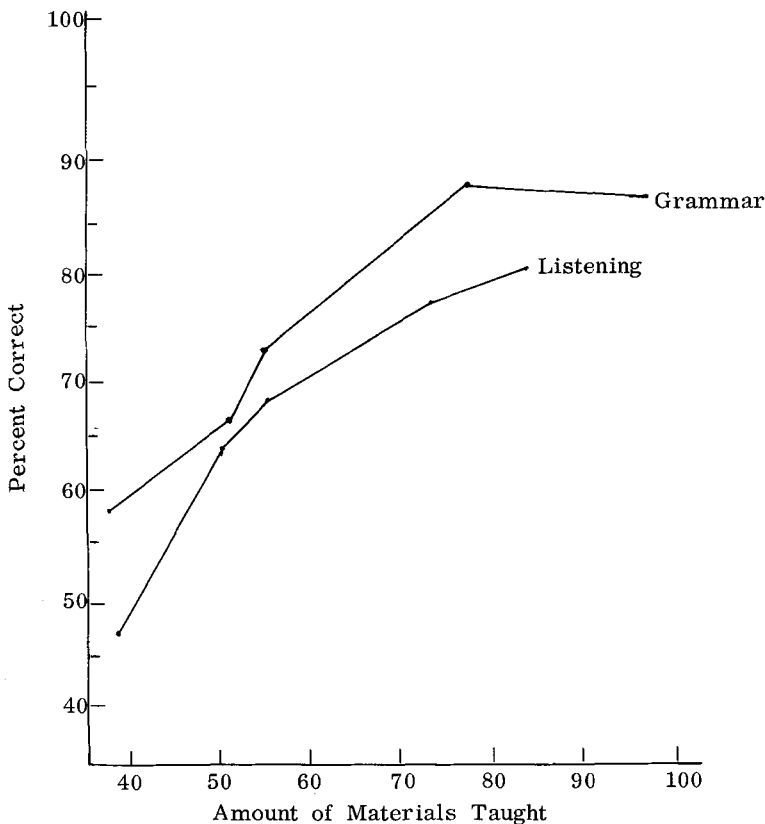


FIG. 1. Total Test Performance of E.F.L. Classes

however, because amount of material taught and test performance are both also correlated with initial English language proficiency.

All classes performed better on test items based upon text materials they had been formally taught than upon items based upon untaught materials. In general, classes which had been taught more performed better on tests of material taught than did classes which had been taught less. Also, performance on tests of untaught material was positively related to amount of material taught, and to performance on tests of taught material. Figure 2 presents the percentages of correctly answered test items based on text materials formally taught and not formally taught as a function of the amount of material taught.

These data do not provide an answer to the question of the possibility of sequential mastery of foreign language course materials. They do, however, suggest that the degree of mastery of a small amount of foreign language grammatical material will be less than the degree of mastery of a larger amount.

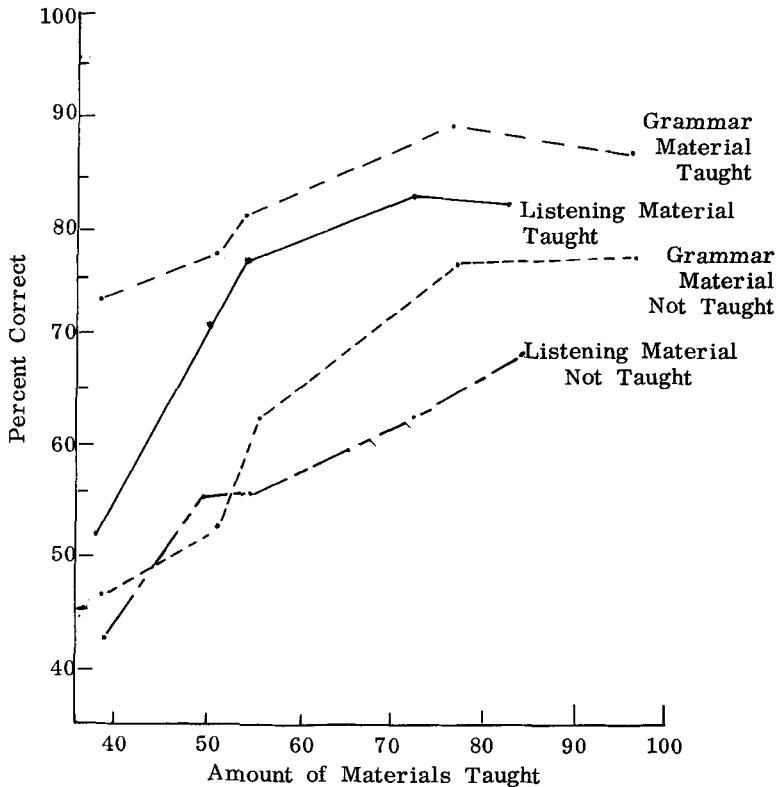


FIG. 2. Performance of E.F.L. Classes on Test Items Covering Text Materials Taught and Text Materials Not Taught

Discussion

The presentation of "negative results" of experiments is a risky business at best. The two examples of *post hoc* arguments which follow illustrate why strict interpretation of negative results is not possible:

1. The subjects in Experiment I are highly selected; they are likely so motivated and have such high verbal abilities that they will learn a foreign language without formal assistance; their talents for teaching themselves are as great as the talents of their instructors for teaching them; quite different results might be expected in an experiment using less exceptional subjects.
2. The number of subjects in Experiments I and III was so small that real differences between the populations from which they were drawn are extremely difficult to find.

Yet despite the fact that these experiments do not provide unequivocal answers to the questions posed in the beginning of this paper, the findings, negative though they are, do suggest points which merit discussion.

The results of Experiment I suggest that foreign language teachers may not be right in some of their assumptions about the values of formal foreign language instruction. Wolfe has stated that, in general, the child learns a foreign language from his environment better than in the classroom whereas the adult learns better in the classroom.⁸ This statement may be true for certain classes of adult learners, but it was not demonstrably true for the subjects in Experiment I. Perhaps Wolfe's "immigrant in an alien speech community" who acquires "only a 'broken,' imperfect fluency" has learned the quantity and/or variety of English which is functional within the sociolinguistic settings in which he normally operates, but his language abilities would be inadequate in certain social situations as determined by value clusters, domains, settings, role relationships and interaction types which are in fact quite remote from his experience.⁹

⁸David L. Wolfe, "Some Theoretical Aspects of Language Learning and Language Teaching" *Language Learning* XVII (1967), 3 & 4, p. 174.

⁹For a summary of a socio linguistic model incorporating these variables see Robert L. Cooper "How Can We Measure the Roles which a Bilingual's Language Play in his Everyday Behavior?" In the Preprints of *The Description and Measurement of Bilingualism*, Toronto, Canadian National Commission for UNESCO, 1967.

In the course of their law studies the subjects in Experiment I were exposed to, and required to use, a variety of English which might be termed "academic." The situation and language variety of the English classes must also have been largely "academic." The experiment showed no advantage to receiving daily two hours of English instruction rather than one hour, or even to receiving some instruction rather than none.

The primary value of foreign language classes for adults may be, therefore, the creation of a range of situations in which the student may learn the language varieties appropriate to those situations and not the provision of drill to produce within the student so-called automatic habits corresponding to the so-called basic patterns of the foreign language. This conclusion is admittedly a very long leap from the findings of Experiment I. But it is an interpretation of those data which is also consistent with results of recent research on language acquisition,¹⁰ with the low correlations typically found between language test scores and grade point averages of foreign students in U.S. universities.¹¹ and with the different relationships found between English proficiency test scores and foreign student grades in courses in which language variety and sociolinguistic settings differ.¹² Carroll has noted that:

When conducted in a school situation, second-language learning tends to be largely. . . *formal* learning—learning guided by conscious, deliberate effort on the part of the learner; there is also considerable infusion of. . . *technical* learning—learning guided by the application of rules and logic. Very little of it is similar to the kind of *informal* learning. . . that occurs in much early first-language learning. Although formal and technical learning may have some place in second-language learning, it is probable that a faster, more appropriate kind of learning can be attained by shifting the balance in favor of "informal" learning.¹³

The results of Experiment I do not contradict this admonition. Similarly the contention of Halliday, McIntosh and Strevens that teaching in the language is more effective than the more usual sort of for-

¹⁰See especially Leon A. Jakobovits' discussion of the role of practice and teaching habit integration and automaticity in "Implications of Recent Psycholinguistic Developments for the Teaching of a Second Language." *Language Learning*, this issue, pp. 89-109.

¹¹Richard E. Spencer, "Academic Achievement of Foreign Students 1960-1967." Research Report No. 259, University of Illinois, Office of Instructional Resources, Measurement and Research Division, 1967.

¹²J. A. Upshur, "English Language Tests and Prediction of Academic Success." Paper presented to the National Association for Foreign Student Affairs, April, 1967.

¹³John B. Carroll, *Language and Thought*, Englewood Cliffs, New Jersey, Prentice-Hall, Inc., 1964, p. 43.

eign language class¹⁴ receives support even though it is not proved.

Experiment II establishes the fact that significant learning of the materials taught in the initial portion of a language course does occur. Experiment III examines the effects of this learning upon subsequent learning. No measurable effects were found. An explanation for this apparent anomaly can best be found by reference to Experiment IV. An extrapolation of the results of that experiment can provide the basis for an explanation.

Within the course, the more materials different groups had been presented the greater was their mastery—as measured by the achievement tests used—of those materials. In the absence of evidence to the contrary, one might infer that a single group receiving foreign language instruction would exhibit the same kind of learning curve. That is, one might infer that in Experiment IV the positive relation between amount of material taught in the course and mastery of that material is not entirely attributable to pre-experimental differences in the knowledge of English, that if one group of subjects were to be tested for mastery of presented materials at a number of times throughout a course, the same positive relation between amount of materials taught and mastery of those materials would obtain.¹⁵

At the beginning of their course the subjects of Experiment III had not completely mastered the initial materials. The simplest explanation of this fact would be that these materials had not previously been taught. The inference based on the results of Experiment IV suggest a different interpretation, however: the subjects of Experiment III had received prior instruction on the initial materials; in fact, because their placement test scores were as high as the post-test scores of the subjects in Experiment II, one would conclude that they had been taught a great deal more than the initial materials. Thus the Experiment III control group received “re-teaching” on the first ten lessons of the texts, and this “re-teaching was essentially a waste of time. The fact that the subjects of Experiment II did show improvement in mastery of the initial materials is irrelevant to this interpretation.

Experiment IV suggests a form of learning language systems which bears further investigation. (In the following discussion an assumption is made that teaching is positively related to learning. Although this assumption did not prove tenable in Experiment I, it

¹⁴M. A. K. Halliday, Angus McIntosh and Peter Strevens, *The Linguistic Sciences and Language Teaching*, London, Longmans, Green and Co., Ltd., 1964, pp. 252-253.

¹⁵Unfortunately, a large set of parallel, internal criterion tests is not yet available, so the validity of this assumption has not been investigated.

is more reasonable in the case of an intensive course in which a large share of the student's foreign language experience is governed by his course and takes place in the classroom.) In the early stages of learning, elements (or dependent sub-systems) are imperfectly learned; those elements which have been specifically presented are learned somewhat better than elements which have not. It is perhaps the case that an external structure and external significance (e.g., translation equivalents) is being learned at this point.¹⁶ As learning progresses the degree of learning of the total of the specifically presented elements at first increases more rapidly and then less rapidly than the degree of learning of the elements which have not been presented. This may be the result of the learner's perceiving the internal structure of presented elements and somewhat more slowly inferring structure for the larger system. Figure 3 illustrates the hypothesized performance curves for presented and unpresented elements in a closed language system.

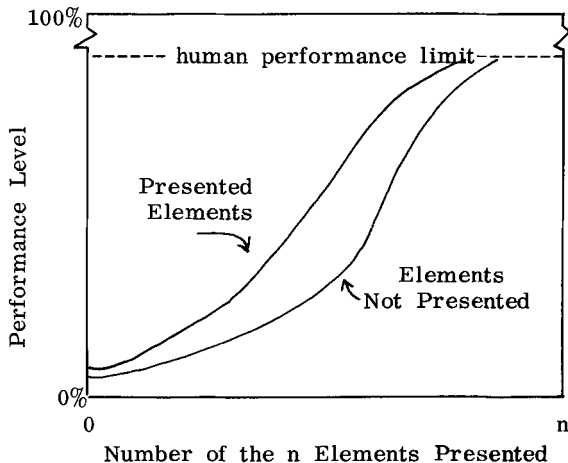


FIG. 3. Hypothesized Learning Curves for a Closed Language System of n Elements

Summary

Four experiments were conducted to answer three related questions about language learning and teaching. Although definite answers were not provided, the following conclusions were suggested:

¹⁶Cf. Wendell R. Garner, *Uncertainty and Structure as Psychological Concepts*, New York, John Wiley and Sons, Inc., 1962, pp. 138-174.

1. Foreign language courses may at this time be less effective means for producing language learning (at least for some learners) than the use of the language in other activities.
2. The structure underlying a sub-set of sentences of a language is not completely learned through the presentation and drill of that sub-set of sentences, but what structure is learned is appropriate in unanticipated ways to performance on other sentences.
3. From conclusion 2 immediately above, it is not reasonable to expect foreign language learners to produce and interpret correctly each of the sentences of the language in the order that they are presented and as soon as they have been drilled.
4. Foreign language achievement tests for placement or diagnosis which utilize sentences of the language (i.e., which do not test directly knowledge of the underlying structure) can not be expected to indicate which "sentence types" have been taught and not taught. From conclusions 2 and 3 above, a knowledge of which sentence types have been taught may in fact be of little value to one who wants to know how much of the structure of a language has been learned.