*Journal of Experimental Psychology*
1972, Vol. 96, No. 1, 124–129

# OPTIMIZING THE LEARNING OF A SECOND-LANGUAGE VOCABULARY

RICHARD C. ATKINSON [1]

*Stanford University*

The problem is to optimize the learning of a large German–English vocabulary. Four optimization strategies are proposed and evaluated experimentally. The first strategy involves presenting items in a random order and serves as a benchmark against which the others can be evaluated. The second strategy permits *S* to determine on each trial of the experiment which item is to be presented, thus placing instruction under "learner control." The third and fourth strategies are based on a mathematical model of the learning process; these strategies are computer controlled and take account of *S*'s response history in making decisions about which items to present next. Performance on a delayed test administered 1 wk. after the instructional session indicated that the learner-controlled strategy yielded a gain of 53% when compared to the random procedure, whereas the best of the two computer-controlled strategies yielded a gain of 108%. Implications of the work for a theory of instruction are considered.

This article examines the problem of individualizing the instructional sequence so that the learning of a second-language vocabulary occurs at a maximum rate. The constraints imposed on the experimental task are those that typically apply to vocabulary learning in an instructional laboratory. A large set of German–English items are to be learned during an instructional session which involves a series of discrete trials. On each trial, one of the German words is presented and *S* attempts to give the English translation; the correct translation is then presented for a brief study period. A predetermined number of trials is allocated for the instructional session, and after some intervening period of time a test is administered over the entire vocabulary set. The problem is to specify a strategy for presenting items during the instructional session so that performance on the delayed test will be maximized. The instructional strategy will be referred to as an *adaptive teaching system* to the extent that it takes into account *S*'s response history in deciding which items to present from trial to trial.

In this study four strategies for sequencing the instructional material are considered. One strategy (designated RO) is to cycle through the set of items in a random order; this strategy is not expected to be particularly effective, but it provides a benchmark against which to evaluate other procedures. A second strategy (designated SS) is to let *S* determine for himself how best to sequence the material. In this mode, *S* decides on each trial which item is to be tested and studied; the learner rather than an external controller determines the sequence of instruction.

The third and fourth sequencing schemes are based on a decision-theoretic analysis of the instructional task (Atkinson, 1972). A mathematical model of learning that has provided an accurate account of vocabulary acquisition in other experiments is assumed to hold in the present situation. This model is used to compute, on a trial-by-trial basis, an individual *S*'s current state of learning. Based on these computations, items are selected for test and study so as to optimize the level of learning achieved at the termination of the instructional session. Two optimization strategies derived from this type of analysis are examined in the present experiment. In one case, the computations for determining an optimal strategy are carried out assuming that all vocabulary items are of equal difficulty; this strategy is designated OE (i.e., optimal under the assumption of equal item difficulty). In the other case, the

[1] Requests for reprints should be sent to Richard C. Atkinson, Department of Psychology, Stanford University, Stanford, California 94305.

computations take into account variations in difficulty level among items; this strategy is called OU (i.e., optimal under the assumption of unequal item difficulty). The details of these two strategies are complicated and can be discussed more meaningfully after the experimental procedure and results have been presented. Both represent adaptive teaching strategies in the sense defined above, because the item presented to *S* on each trial is determined by his history of responses up to that trial.

The concern of the experiment is to evaluate the relative effectiveness of the four instructional strategies. Of particular interest is whether strategies derived from a theoretical analysis of the learning process can be as effective as a procedure where *S* makes his own decisions.

## Method

*Subjects.*—The *S*s were 120 undergraduates enrolled at Stanford University; 30 *S*s were randomly assigned to each of the four experimental groups. None of the students had prior course work in German and none professed familiarity with the language. The *S*s were run in groups of 8, with 2 *S*s in each group assigned to one of the four experimental conditions.

*Materials.*—The instructional material involved 84 German–English items; all words were concrete nouns typically taught during the first course in German. Seven display lists of 12 German words each were formed that were judged to be of roughly equal difficulty. A display list involved a vertical array of German words numbered from 1 to 12; the seven lists were arranged in a round robin and were always cycled through in the same order.

*Procedure.*—The *S*s were required to participate in two sessions: an *instructional session* that lasted approximately 2 hr. and a much shorter *test session* administered 1 wk. later. The test session was the same for all *S*s and involved a test over the entire set of items. The four experimental groups were distinguished by the sequencing strategy (RO, SS, OE, OU) employed during the instructional session. The experiment was conducted in the Computer-Based Learning Laboratory at Stanford University. The control functions were performed by programs run on a modified PDP-1 computer (Digital Equipment Corp.) operating under a time-sharing system. Eight teletypewriters were housed in a soundproof room and faced a projection screen mounted on the front wall.

The instructional session involved a series of discrete trials. Each trial was initiated by projecting one of the display lists on the front wall of the room; the list remained on the screen throughout the trial. The *S*s were permitted to inspect the list for approximately 10 sec. In the RO, OE, and OU conditions this inspection period was followed by the computer typing a number from 1 to 12 on each *S*'s teletypewriter indicating the item to be tested on that trial; the number typed on a given teletypewriter depended on that *S*'s particular control program. In the SS condition, *S* typed 1 of 12 numbered keys during the inspection period to indicate to the computer which item he wanted to be tested on. At the end of the inspection period, *S* was required to type out the English translation for the designated German word and then strike the "slash" key, or if unable to provide a translation to simply hit the "slash" key. After the "slash" key had been activated the computer typed out the correct translation and spaced down two lines in preparation for the next trial. The trial terminated with the offset of the display list and the next trial began immediately with the onset of the next display list in the round robin of lists. A complete trial took approximately 20 sec. and the timing of events (within and between trials) was synchronous for the eight *S*s run together. The instructional session involved 336 trials (with a 5-min. break in the middle), which meant that each display list was presented 48 times. In the RO condition, this number of trials permitted each of the items on a list to be tested and studied an average of 4 times.

The delayed-test session, conducted 7 to 8 days later, was precisely the same for all *S*s. All testing was done on the teletypewriters. A trial began with the computer typing a German word, and *S* was then required to type the English translation; *S* received no feedback on the correctness of the response. The 84 German items were presented in a different random order for each *S*. During the delayed-test session the trial sequence was self-paced.

All *S*s were told at the beginning of the experiment that there would be a delayed-test session and that their principal goal was to achieve as high a score as possible on that test. They were told, however, not to think about the experiment or rehearse any of the material during the intervening week; these instructions were emphasized at the beginning and end of the instructional session and later reports from *S*s confirmed that they made no special effort to rehearse the material during the week between instruction and the delayed test. The instructions emphasized that *S* should try to provide a translation for every item tested during the instructional session; if *S* was uncertain but could offer a guess he was encouraged to enter it. In the RO, OE, and OU conditions, no additional instructions were given. In the SS condition, *S*s were told that their trial-to-trial selection of items should be done with the aim of mastering the total set of vocabulary items. They were instructed that it was best to test and study on words they did not know rather than on ones already mastered.

## Results

The results of the experiment are summarized in Fig. 1. On the left side of the

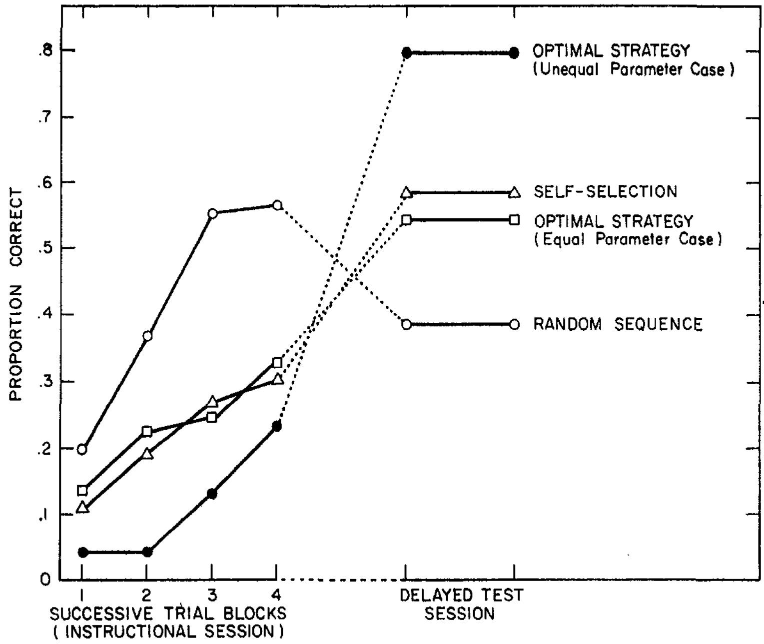


FIG. 1. Proportion of correct responses in successive trial blocks during the instructional session and on the delayed test administered 1 wk. later.

figure, data are presented for performance during the instructional session; on the right side are results from the delayed test. The data from the instructional session are presented in four successive blocks of 84 trials each; for the RO condition this means that on the average each item was presented once in each of these blocks. Note that performance during the instructional session is best for the RO condition, next best for the OE condition which is slightly better than the SS condition, and poorest for the OU condition; these differences are highly significant, $F$ (3, 116) = 21.3, $p < .001$. The order of the experimental groups on the delayed test is completely reversed. The OU condition is by far best with a correct response probability of .79; the SS condition is next with .58; the OE condition follows closely at .54; and the RO condition is poorest at .38, $F$ (3, 116) = 18.4, $p <$ .001. The observed pattern of results is what one would expect. In the SS condition, *S*s are trying to test themselves on items they do not know; consequently, during the instructional session, they should have a lower proportion of correct responses than *S*s run on the RO procedure where items are tested at random. Similarly, the OE and OU conditions involve a procedure that attempts to identify and test those items that have not yet been mastered and also should produce high error rates during the instructional session. The ordering of groups on the delayed test is reversed since now all words are tested in a nonselective fashion; under these conditions the proportion of correct responses provides a measure of *S*'s mastery of the total set of vocabulary items.

The magnitude of the effects observed on the delayed test are large and of practical significance. The SS condition (when compared to the RO condition) leads to a relative gain of 53%, whereas the OU condition yields a relative gain of 108%. It is interesting that *S* can be very effective in determining an optimal study sequence, but not so effective as the best of the two adaptive teaching systems.

## Discussion

At this point, we turn to an account of the theory on which the OU and OE schemes are based. Both schemes assume that acquisition of a second-language vocabulary can be described by a fairly simple learning model. It is postulated that a given item is in one of three states (P, T, and U) at any moment in time. If the item is in State P, then its translation is known and this knowledge is "relatively" permanent in the sense that the learning of other items will not interfere with it. If the item is in State T, then it is also known but on a "temporary" basis; in State T the learning of other items can give rise to interference effects that cause the item to be forgotten. In State U the item is not known, and $S$ is unable to give a translation. Thus in States P and T a correct translation is given with probability one, whereas in State U the probability is zero.

When Item $i$ is presented for test and study the following transition matrix describes the possible change in state from the onset of the trial to its termination:

$$\mathbf{A}_i = \begin{array}{c} \\ \mathrm{P} \\ \mathrm{T} \\ \mathrm{U} \end{array}\!\!\begin{array}{c} \begin{array}{ccc} \mathrm{P} & \mathrm{T} & \mathrm{U} \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ x_i & 1 - x_i & 0 \\ y_i & z_i & 1 - y_i - z_i \end{bmatrix} \end{array}.$$

Rows of the matrix represent the state of Item $i$ at the start of the trial and columns the state at the end of the trial. On a trial when some item *other* than Item $i$ is presented for test and study (whether that item is a member of Item $i$'s display list or some other display list), transitions in the learning state of Item $i$ also may take place. Such transitions can occur only if $S$ makes an error on the trial; in that case the transition matrix applied to Item $i$ is as follows:

$$\mathbf{F}_i = \begin{array}{c} \\ \mathrm{P} \\ \mathrm{T} \\ \mathrm{U} \end{array}\!\!\begin{array}{c} \begin{array}{ccc} \mathrm{P} & \mathrm{T} & \mathrm{U} \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 - f_i & f_i \\ 0 & 0 & 1 \end{bmatrix} \end{array}.$$

Basically, the idea is that when some other item is presented to which $S$ makes an error (i.e., an item in State U), then forgetting may occur for Item $i$ if it is in State T.

To summarize, when Item $i$ is presented for test and study transition Matrix $\mathbf{A}_i$ is applied; when some other item is presented that elicits an error then Matrix $\mathbf{F}_i$ is applied. The above assumptions provide a complete account of the learning process. For the task considered in this article it is also assumed that Item $i$ is either in State P (with probability $g_i$) or in State U (with probability $1 - g_i$) at the start of the instructional session; $S$ either knows the correct translation without having studied the item or does not. The parameter vector $\phi_i = [x_i, y_i, z_i, f_i, g_i]$ characterizes the learning of a given Item $i$ in the vocabulary set. The first three parameters govern the acquisition process; the next parameter, forgetting; and the last, $S$'s knowledge prior to entering the experiment.

For a more detailed account of the model the reader is referred to Atkinson and Crothers (1964) and Calfee and Atkinson (1965). It has been shown in a series of experiments that the model provides a fairly good account of vocabulary learning and for this reason it was selected to develop an optimal procedure for controlling instruction. We now turn to a discussion of how the OE and OU procedures were derived from the model. Prior to conducting the experiment reported in this article, a pilot study was run using the same word lists and the RO procedure described above. Data from the pilot study were employed to estimate the parameters of the model; the estimates were obtained using the minimum chi-square procedures described in Atkinson and Crothers. Two separate estimates of parameters were made. In one case it was assumed that the items were equally difficult, and data from all 84 items were lumped together to obtain a single estimate of the parameter vector $\phi$; this estimation procedure will be called the equal-parameter case (E case), since all items are assumed to be of equal difficulty. In the second case data were separated by items, and an estimate of $\phi_i$ was made for each of the 84 items (i.e., $84 \times 5 = 420$ parameters were estimated); this procedure will be called the unequal-parameter case (U case). In both the U and E cases, it was assumed that there were no differences among $S$s; this homogeneity assumption regarding learners will be commented upon later. The two sets of parameter estimates were used to generate the optimization schemes previously referred to as the OE and OU procedures; the former based on estimates from Case E and the latter from Case U.

In order to formulate an instructional strategy, it is necessary to be precise about the quantity to be maximized. For the present experiment the goal is to maximize the total number of items $S$ correctly translates on the delayed test.[2] To do this, we need to specify the theoretical relationship between the state of learning at the end of the instructional session and performance on the delayed test. The assumption made here is that only those items in State P at the end of the instructional session will be translated correctly on the delayed test; an item in State T is presumed to be forgotten during the intervening week. Thus, the problem of maximizing delayed-test performance involves, at least in theory, maximizing the number of items in State P at the termination of the instructional session.

Having numerical values for parameters and knowing *S*'s response history, it is possible to estimate his current state of learning.[3] Stated more precisely, the learning model can be used to derive equations and, in turn, compute the probabilities of being in States P, T, and U for each item at the start of Trial $n$, conditionalized on *S*'s response history up to and including Trial $n - 1$. Given numerical estimates of these probabilities, a strategy for optimizing performance is to select that item for presentation (from the current display list) that has the greatest probability of moving into State P if it is tested and studied on the trial. This strategy has been termed the one-stage optimization procedure because it looks ahead only one trial in making decisions. The true optimal policy (i.e., an $N$-stage procedure) would consider all possible item-response sequences for the remaining trials and select the next item so as to maximize the number of items in State P at the termination of the instructional session. Unfortunately, for the present case the $N$-stage policy cannot be applied because the necessary computations are too time consuming even for a large-scale computer. A series of Monte Carlo studies indicates that the one-stage policy is a good approximation to the optimal strategy for a variety of Markov learning models; it was for this reason, as well as the relative ease of computing, that the one-stage procedure was employed. For a discussion of one-stage and $N$-stage policies and Monte Carlo studies comparing them see Groen and Atkinson (1966), Calfee (1970), and Laubsch (1970).

The optimization procedure described above was implemented on the computer and permitted decisions to be made on-line for each *S* on a trial-by-trial basis. For *S*s in the OE group, the computations were carried out using the five parameter values estimated under the assumption of homogeneous items (E-case); for *S*s in the OU group the computations were based on the 420 parameter values estimated under the assumption of heterogeneous items (U-case).

The OU procedure is sensitive to interitem differences and consequently generates a more effective optimization strategy than the OE procedure. The OE procedure, however, is almost as effective as having *S* make his own instructional decisions and far superior to a random presentation scheme. If individual differences among *S*s also are taken into account, then further improvements in delayed-test performance should be possible; this issue and methods for dealing with individual differences are discussed in Atkinson and Paulson (1972).

The study reported here illustrates one approach that can contribute to the development of a theory of instruction (Hilgard, 1964). This is not to suggest that the OU procedure represents a final solution to the problem of optimal item selection. The model upon which this strategy is based ignores several important factors, such as interitem relationships, motivation, and short-term memory effects (Atkinson & Shiffrin, 1968, p. 190). Undoubtedly, strategies based on learning models that take these variables into account would yield superior procedures.

Although the task considered in this article deals with a limited form of instruction, there are at least two practical reasons for studying

[2] Other measures can be used to assess the benefits of an instructional strategy; e.g., in this case weights could be assigned to items measuring their relative importance. Also costs may be associated with the various actions taken during an instructional session. Thus, for the general case, the optimization problem involves assessing costs and benefits and finding a strategy that maximizes an appropriate function defined on them. For a discussion of this issue see Atkinson and Paulson (1972), Dear, Silberman, Estavan, and Atkinson (1967), and Smallwood (1971).

[3] The *S*'s *response history* is a record for each trial of the vocabulary item presented and the response that occurred. It can be shown that there exists a *sufficient history* that contains only the information necessary to estimate *S*'s current state of learning; the sufficient history is always a function of the complete history and the assumed learning model (Groen & Atkinson, 1966). For the model considered in this paper the sufficient history is fairly simple. It is specified in terms of individual vocabulary items for each *S*; we need to know the ordered sequence of correct and incorrect responses to a given item plus the number of errors (to other items) that intervene between each presentation of the item.

it. First, this type of task occurs in numerous learning situations; no matter what the pedagogical orientation, any initial reading program or foreign-language course involves some form of list learning. In this regard it should be noted that a modified version of the OU strategy has been used successfully in the Stanford computer-assisted instruction program in initial reading (Atkinson & Fletcher, 1972). Second, the study of simple tasks that can be understood in detail provides prototypes for analyzing more complex optimization problems. At present, analyses comparable to those reported here cannot be made for many instructional procedures of central interest to educators, but examples of this sort help to clarify the steps involved in devising and testing optimal strategies.[4]

[4] For a review of work on optimizing learning and references to the literature see Atkinson (1972).

## REFERENCES

Atkinson, R. C. Ingredients for a theory of instruction. *American Psychologist*, 1972, **27**, 921–931.

Atkinson, R. C., & Crothers, E. J. A comparison of paired-associate learning models having different acquisitions and retention axioms. *Journal of Mathematical Psychology*, 1964, **2**, 285–315.

Atkinson, R. C., & Fletcher, J. D. Teaching children to read with a computer. *The Reading Teacher*, 1972, **25**, 319–327.

Atkinson, R. C., & Paulson, J. A. An approach to the psychology of instruction. *Psychological Bulletin*, 1972, **78**, 49–61.

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory*. Vol. 2. New York: Academic Press, 1968.

Calfee, R. C. The role of mathematical models in optimizing instruction. *Scientia: Revue Internationale de Synthèse Scientifique*, 1970, **105**, 1–25.

Calfee, R. C., & Atkinson, R. C. Paired-associate models and the effect of list length. *Journal of Mathematical Psychology*, 1965, **2**, 255–267.

Dear, R. E., Silberman, H. F., Estavan, D. P., & Atkinson, R. C. An optimal strategy for the presentation of paired-associate items. *Behavioral Science*, 1967, **12**, 1–13.

Groen, G. J., & Atkinson, R. C. Models for optimizing the learning process. *Psychological Bulletin*, 1966, **66**, 309–320.

Hilgard, E. R. (Ed.) *Theories of learning and instruction: The sixty-third yearbook of the National Society for the Study of Education.* Chicago: University of Chicago Press, 1964.

Laubsch, J. H. Optimal item allocation in computer-assisted instruction. *IAG Journal*, 1970, **3**, 295–311.

Smallwood, R. D. The analysis of economic teaching strategies for a simple learning model. *Journal of Mathematical Psychology*, 1971, **8**, 285–301.

(Received February 4, 1972)