Prediction of complex verbal learning¹

E. O. EIMER AND R. J. SENTER

UNIVERSITY OF CINCINNATI

Estes' (1959) linear stimulus sampling model was modified to predict response probability on trial n on the basis of response probability on trial 1. The model was employed to predict three learning functions for a group of Ss learning a poem. Statistical analysis indicated high predictive power. The theoretical functions, based on observation of performance after 5 min. of practice accounted for 99% of the trials variance over 35 min. of performance. The findings are discussed in terms of applied and theoretical considerations.

Mathematical models are at times presented with the implication that they apply to complex verbal learning (Estes, 1964). Actual evidence for such application is scarce, for applications of these models are usually restricted to two-choice situations in which the population of stimuli is constant and all stimuli have an equal probability of being sampled on the first trial. In the present study, a modification of Estes' linear stimulus sampling model (1959) was employed as a predictor of poetry learning.

The linear stimulus sampling model theoretically falls far short of applicability to complex verbal learning. Despite its restrictions, however, it has two advantages over most other models. One of these advantages is that it is simple. In the function

$$p_{n+1} = (1 - \theta) p_n + \theta \quad (1)$$

the probability of a response on trial $n + 1$ is expressed in terms of response probability on trial n , and in terms of θ , a parameter with a value between 0 and 1.

The second advantage is that the model seems rugged enough to stand various violations of its assumptions and yet to retain predictive power. This point is illustrated in its present application.

The linear stimulus sampling model can be simplified further if the assumption is made that the response probability p_0 of a hypothetical trial No. 0 equals 0. From this assumption it would follow that $p_1 = \theta$ and

$$p_n = 1 - (1 - p_1)^n \quad (2)$$

This equation permits prediction of performance on trial n on the basis of p_1 only.

Method

As a part of a larger experiment, 68 students enrolled in an introductory psychology course at the University of Cincinnati were given the task of learning a 64-word selection from E. A. Poe's *Tamerlain*. All Ss were initially unfamiliar with the poem. A test booklet consisting of 25 identical sheets with the poem written on the upper half was provided. On the lower half of each sheet underlined spaces proportional to the lengths of the words comprising the poem were presented. The upper half of each sheet was covered by an opaque overlay

with a space for writing a "time number." During the administration, the experimenter wrote consecutive numbers on the blackboard, starting with 1, and changing them to the next higher integer every 15 sec. The Ss were instructed to read the poem on the first page, then to go on to the next page, record the time number that was presently on the blackboard, write down in the spaces provided as much of the poem as they could remember, flip up the overlay sheet to review their reproduction against the original, then go on to the next page, write down the time number, etc. Ss were self paced. This process was repeated for 37.5 min.

Results

Performance was evaluated in terms of number of correct words in each of seven reproductions, i.e., those on which Ss were working at the end of 5, 10, 15, 20, 25, 30, and 35 min. from the starting time. The trials effect was significant ($F = 350.69$, $df = 6/402$). On the basis of the first trial p_1 was established. A trend analysis was performed, testing the observed data against the trend expressed by the function

$$p_n = 1 - .54^n$$

which was obtained from equation (2) and p_1 . The trend accounted for .991 of the trials' variance (see Fig. 1).

In order to determine whether equation (2) also had predictive power for extreme cases, two groups of 10 Ss each were selected from the original sample. One of these groups consisted of the 10 Ss with highest overall performance; the other was composed of the 10 Ss with lowest overall performance. A separate analysis of variance was performed for each group. The trials effect was significant in each case (high: $F = 88.74$, $df = 6/54$; low: $F = 63.13$, $df = 6/54$). For each group, p_1 was established, and the observed data were tested against the resulting functions, based on equation (2). The trends accounted for .988 and .992 of the trials variances for high and low groups, respectively (see Fig. 2).

Discussion

From the viewpoint of applied psychology, the present findings suggest that the model as described in equation (2) yields an efficient predictor of later performance based on the observation of initial performance only. Performance after 35 min. can be predicted rather accurately after 5 min. (See Fig. 2).

From a theoretical viewpoint the findings are, in some respects, puzzling. Learning a poem is a complex task. It is doubtful that one word represents one stimulus, or that the sampling of one word is independent of the sampling of another word. There is no reason to believe that once a word has been correctly recalled, it will

continue to be correctly recalled on all following trials. (In fact, examination of the reproductions indicates the contrary.) Finally, the assumption that $p_0 = 0$ may be seriously questioned. In summary, all assumptions of the model were violated to a greater or lesser extent, and all that may be said for the application of this model to the present data is that it predicts them well.

Inspection of Fig. 1 reveals that the deviations of the data points from the function, even though they are small, are systematic. If p_0 is assumed to be .057 rather than 0, the resulting function describes the data better. This procedure, however, entails determining an additional parameter, p_0 , which obviates performance prediction on the basis of p_1 only.

While the present research may be taken to support Estes' linear stimulus sampling model, and while an efficient performance predictor has been suggested, it is also patently evident (Fig. 1) that the model, particularly as interpreted in equation (2), is not the

Fig. 1. Percent correct responses, observed and theoretical, as a function of time: group mean function. ($N = 68$).

Fig. 2. Percent correct responses, observed and theoretical, as a function of time: extreme subgroup mean functions. ($N_{\text{high}} = 10$; $N_{\text{low}} = 10$).

best possible predictor of complex verbal learning. Until further research is done, however, it seems to be one of the better *existing* predictors, if only by virtue of the fact that others apparently have not been employed with more success. Finally, the present research indicates that prediction of complex verbal learning does not lie beyond the scope of existing mathematical models.

References

- Estes, W. K. Probability learning. In A. W. Melton (Ed.), *Categories of human learning*. New York: Academic Press, 1964. Pp. 89-128.
- Estes, W. K., & Suppes, P. Foundations of linear models. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford: Stanford University Press, 1959. Pp. 137-179.

Note

1. This study was supported by Air Force Contract No. AF 33-(615)1046.