Behavior Research Methods & Instrumentation
1983, 15 (6), 577-579

Knowledge of the alphabet: A comparison between letter counts and subjective reports

PHILIP H. MARSHALL and CLAY E. GEORGE
Texas Tech University, Lubbock, Texas

Subjects' estimates of the frequency of occurrence of the letters of the alphabet were compared with previously reported letter-frequency counts. The results indicated an overall good relationship between actual and judged rank, but there were consistent letter-specific under- and overestimations. These inaccuracies were not accounted for by letter versatility, first-position frequency, or order in the alphabet. There was evidence that subject-derived estimates of letter frequency were somewhat better predictors of reaction time performance on letter-processing tasks.

Computerized counts of letters, words, and certain orthographic combinations are of interest because of increased research in letter processing, reading, and word recognition and storage (Rubin, 1980; Solso, Juel, & Rubin, 1982; Solso & King, 1976). Solso et al (1982) reviewed these efforts. Our interest addressed the extent to which people's knowledge or subjective estimates of letter frequency match the results of these computer counts. To the extent that individuals have similar information, these norms, scales, and counts can be expected to have explanatory or predictive potential for explaining letter-processing tasks.

The utility of one of these counts was reported by Appelman and Mayzner (1981), who examined the relationship between the simple measure of letter frequency and the performance measures of letter recognizability and letter discrimination. The specific scale of letter frequency was determined by Solso and King (1976). Rank-order correlations were computed between the letters' ranks and the ranks on various letter-performance tasks (58 studies) conducted over a period of nearly 100 years. The results indicated that letter frequency has no effect on letter-recognition tasks in which subjects merely report letters, but that letter frequency does positively correlate with various reaction time measures of letter discrimination. Because these correlations were not especially large, a secondary focus of the present study was to determine whether subject-derived ranks would yield better predictors of the letter-frequency effects.

Attneave (1953) studied the correspondence between actual and psychological or subjective probabilities, and chose knowledge of the letters of the alphabet as the task. Letters represented a stable experiential dimension that reasonably could be assumed to have asymptoted by adulthood. In the condition in Attneave's study that was most similar to the present investigation, subjects were shown, but did not study, a sheet of printed material containing approximately 1,000 letters and were asked to estimate the frequencies of occurrence of all of the letters of the alphabet. The estimates of 90 subjects were compared with those obtained from a sample of slightly more than 10,000 letters from newspaper and magazine sources. The data yielded a correlation of $r = .81$ between the actual and judged frequencies. Attneave pointed out, however, that the correlation merely indicated a lawful relationship and that there were systematic errors in judgment.

The studies reported here involved a partial reappraisal of the Attneave (1953) data and addressed the following questions: How well do people actually know the frequency of occurrence of the letters of the alphabet? If there are discrepancies between the actual and the judged frequencies, are they related in general to errors of the extremes of frequency, or are they of a more specific nature? And finally, are subject-derived or frequency-based ranks more predictive of performance measures?

METHOD

Two studies were conducted to answer these questions. Each required that subjects make frequency judgments of the letters of the alphabet by determining ranks reflecting their relative frequencies in printed material. The studies differed in the way that rank information was obtained.

Study 1

Subjects. The subjects were 30 native English-speaking adult volunteers (15 males and 15 females) recruited from the academic community at Texas Tech University. The age range was from 19 to 70 years, with a mean of 28.8 years (median = 25.5 years).

Materials. Each subject received a single sheet of paper with instructions written on the top half and 26 spaces, each labeled with a single letter of the alphabet, on the bottom half. The letters were in alphabetical order. The subjects were instructed to use each of the numbers 1 through 26 in assigning the ranks,

Thanks are due Gale Wilson for computer programming the frequency analyses. Requests for reprints should be sent to: Philip H. Marshall, Department of Psychology, Texas Tech University, Lubbock, Texas 79409.

with 1 for the most frequent and 26 for the least frequent letter.[1]

Procedure. The subjects were allowed to complete the task at their leisure, with the only restrictions being that it be done in a quiet setting and with no assistance. The task required about 10-15 min to complete.

Study 2

Subjects. The subjects for Study 2 were drawn from the same population as those in Study 1. There were 20 volunteers (14 males and 6 females) whose ages ranged from 18 to 46 years, with a mean of 28.35 years (median = 26.5 years).

Materials. A single sheet, similar to that used in Study 1, was handed to each of the subjects. However, instead of a listing of the alphabet, the sheet contained the numbers 1 through 26, signifying the ranks to be filled. The subjects were asked to assign a letter of the alphabet to each rank.

Procedure. The subjects completed the sheet under the same conditions as in Study 1.

RESULTS

Because the means of the rank judgments from the two studies showed high correspondence and congruence (r = .99), these measures were combined.

Mean ranking judgments for the combined studies (n = 50) are shown in Figure 1. The abscissa shows the ranking of the frequency of the letters of the alphabet as reported by Solso and King (1976). The ordinate represents the mean judged rank for each letter. Also plotted on Figure 1, for comparison purposes, are the rank data from Attneave (1953). To obtain the Attneave ranks, the median frequency estimates reported in his study were rank ordered.

The means of the ranking judgments were in turn rank ordered, and rank-order correlations were computed between the Solso and King (1976) ranks and those now obtained from both the Attneave (1953) and present studies. The correlations were rho = .93

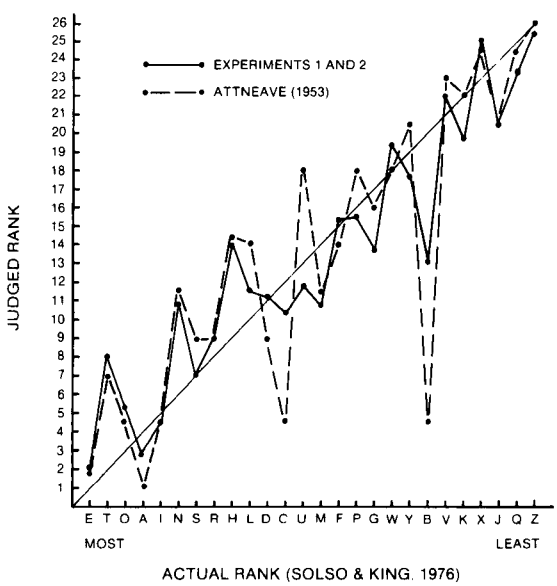

Figure 1. Mean judged ranks for frequency of occurrence of the letters of the alphabet.

Table 1
Rank-Order Correlation Coefficients Between Speed of Correct Responses for Reaction Time Studies and the Solso and King (1976) Frequency Ranks and the Subject-Derived Ranks From the Present Studies

Study	Solso & King (1976)	Present Study
Clement & Carpenter (1970)	.45	.54*
Fox (1975)	.19	.14
Cosky (1976)	.42*	.41*
Cosky (1976)	.57*	.67*
Egeth, Brownell, & Geoffrion (1976)	.36	.41
Podgorny & Garner (1979)	.53*	.58*

**Values needed for significance at p < .05 vary, since not all studies used all 26 letters of the laphabet.*

between the Solso and King (1976) data and the present data, and rho = .84 between the Attneave and the Solso and King data. The rank-order correlation between the Attneave data and the present data was rho = .93.

Inspection of Figure 1 shows that Attneave's (1953) claim that there are general tendencies for underestimation of high-frequency and overestimation of low-frequency letters probably has to be qualified.

We examined the possibility that subjects may have been using other letter attributes to determine letter frequency. [Solso & King (1976) also compiled rank data on letter versatility and first-position (in words) frequency.] The discrepancies between actual letter frequencies and subjects' estimates in the present data may have been due to the subjects' using their knowledge of other dimensions to make their rank judgments. To determine the extent to which this happened, rank-order correlations were computed between letter versatility, first-position frequency (for four-letter words), and order of appearance in the alphabet, with an accuracy measure derived from the present data. The accuracy measure for each letter was taken as the algebraic difference between the Solso and King ranks and the present subject-derived ranks. The correlations were −.28 for versatility, −.16 for first-position frequency, and .30 for order of appearance in the alphabet. None was statistically significant. It appears that the obtained discrepancies are not to be accounted for by subjects' using these other characteristics of the letters.

A second focus of the present study compared the utility of the subject-derived letter ranks with that of the computer count of Solso and King (1976). For this comparison, rank-order correlations were computed between the subject-derived ranks and the data of those studies reported by Appelman and Mayzner (1981), which showed a positive relationship between letter frequency and reaction time for letter-discrimination tasks. For purposes of these correlations, the ranks of the mean ranking judgments were again used. Table 1 shows these comparisons. The overall conclusion is that rank-order correlations tended to be higher for the subject-derived scale in those instances in which correlations were likely.

DISCUSSION

Several conclusions about knowledge of letter frequency may be drawn from the present data and from its comparison with the earlier work of Attneave (1953). First, subjects do have a good appreciation of the relative frequency of occurrence of the letters of the alphabet. Second, the discrepancies from the actual frequencies are not to be attributed to the operation of general ceiling and cellar effects. The systematic deviations serve to caution that letter-count norms may not totally reflect the subjects' perceptions of the frequency of these letters.

Finally, the utility comparisons between the Solso and King (1976) and the subject-derived ranks reported here yielded somewhat higher correlatons for the latter, and this suggests that subjects' knowledge of letter frequency, rather than the actual letter frequencies, may be a more potent determiner of performance in certain letter-processing tasks.

REFERENCES

Appelman, I. B., & Mayzner, M. S. The letter-frequency effect and the generality of familiarity effects on perception. *Perception & Psychophysics*, 1981, **30**, 436-446.

Attneave, F. Psychological probability as a function of experienced frequency. *Journal of Experimental Psychology*, 1953, **46**, 81-86.

Clement, D. E., & Carpenter, J. S. Relative discriminability of visually-presented letter pairs using a same-different choice-reaction time task. *Psychonomic Science*, 1970, **20**, 363-364.

Cosky, M. J. The role of letter recognition in word recognition. *Memory & Cognition*, 1976, **4**, 207-214.

Egeth, H., Brownell, H., & Geoffrion, L. Testing the role of symmetry in letter matching. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, **2**, 429-434.

Fox, J. The use of structural diagnostics in recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, **1**, 57-67.

Podgorny, P., & Garner, W. R. Reaction time as a measure of inter- and intraobject visual similarity: Letters of the alphabet. *Perception & Psychophysics*, 1979, **26**, 37-52.

Rubin, D. C. 51 properties of 125 words: A unit analysis of verbal behavior. *Journal of Verbal Learning and Verbal Behavior*, 1980, **19**, 736-755.

Solso, R. L., Juel, C., & Rubin, D. C. The frequency and versatility of initial and terminal letters in English words. *Journal of Verbal Learning and Verbal Behavior*, 1982, **21**, 220-235.

Solso, R. L., & King, J. F. Frequency and versatility of letters in the English language. *Behavior Research Methods & Instrumentation*, 1976, **8**, 285-286.

NOTE

1. In each study, the subjects were also asked to indicate their confidence in assigning ranks and, additionally, in Study 2, were asked to indicate their order of assigning ranks. Those data have little bearing on the focus of the present studies and will be presented in a subsequent report.

APPENDIX
Solso & King's (1976) and Subject-Derived Ranks of the Letters of the Alphabet

Letter	Solso & King (1976)	Subject-Derived*
A	4	2
B	20	14
C	12	8
D	11	11
E	1	1
F	15	17
G	17	15
H	9	16
I	5	3
J	24	22
K	22	21
L	10	12
M	14	9
N	6	10
O	3	4
P	16	18
Q	25	24
R	8	7
S	7	5
T	2	6
U	13	13
V	21	23
W	18	20
X	23	25
Y	19	19
Z	26	26

**Figure 1 shows the mean ranking judgments given by the 50 subjects. This column gives the ranks of those means so as to be more commensurate with the ranks reported by Solso and King.*

(Manuscript received May 16, 1983;
revision accepted for publication December 2, 1983.)