FREQUENCY, ORTHOGRAPHIC REGULARITY, AND LEXICAL STATUS IN LETTER AND WORD PERCEPTION[1]

Dominic W. Massaro, James E. Jastrzembski, and Peter A. Lucas[2]

UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN

Perceiving and understanding involves a wonderfully adept integration of our immediate environment with prior knowledge and experience. The knowledge of the reader, for example, is as important as, or more important than, the information on the printed page. One compelling issue in reading research is how the reader's higher order knowledge of the lan-

[1]This research was supported by funds from the National Institute of Education to the Wisconsin Research and Development Center for Individualized Schooling, University of Wisconsin, Madison, Wisconsin 53706. Kenneth R. Paap and Sandra L. Newsome made helpful comments on an earlier report of this research.

[2]D. W. Massaro is now affiliated with the University of California, Santa Cruz, California 95064; J. E. Jastrzembski is with Needham, Harper & Steers Advertising, Inc., 300 East Wacker Drive, Chicago, Illinois 60601; and P. L. Lucas with the Department of Psychology, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213. Correspondence concerning this work should be directed to Dominic W. Massaro, Program in Experimental Psychology, University of California, Santa Cruz, California 95064.

guage interacts with lower level perceptual analyses. The specific question addressed in the present article is how the reader's knowledge about orthographic structure is combined with the information derived from visual featural analysis in letter and word recognition. Visual featural analysis refers to the evaluation of the component visual properties of letters leading to letter and word recognition. Orthographic structure refers to the spelling constraints in a written language. Given the considerable amount of predictability in English writing, we ask how the reader utilizes this orthographic structure in word recognition.

I. Language Processing Model

Evaluation of the contributions of visual features and orthographic structure to word recognition can be facilitated by a detailed description of the processes involved in reading. The description we use is part of a more general model of language processing (Massaro, 1975, 1978, 1979a; Massaro, Taylor, Venezky, Jastrzembski, & Lucas, 1980b). According to the model illustrated in Fig. 1, reading can be viewed as a sequence of processing stages. At each stage of processing, memory and process components are represented. Each memory component (indicated by a rectangle) corresponds to the information available at a particular stage of processing. Each process component (indicated by a circle) corresponds to the operations applied to the information held by the memory component. The memory components are temporary storages except for long-term memory, which is relatively permanent. It is assumed that long-term memory supplements the information at some of the processing stages.

During reading, the light pattern reflected from a display of letters is transduced by the visual receptors as the *feature detection* process detects and transmits visual features to preperceptual visual storage (see Fig. 1).

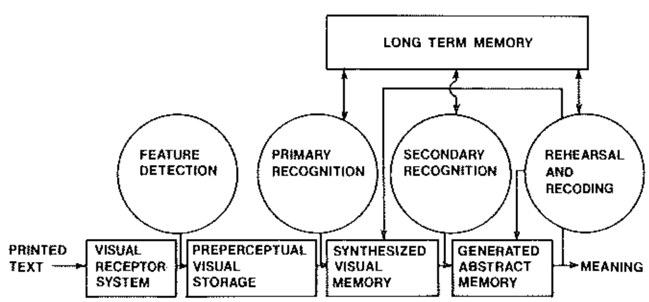

Fig. 1. Schematic representation of stages of processing in reading.

As visual features enter in preperceptual visual storage, the *primary recognition* process attempts to transform these isolated features into a sequence of letters and spaces in synthesized visual memory. To do this, the primary recognition process can utilize information held in long-term memory. For the accomplished reader this includes a list of features for each letter of the alphabet along with information about the orthographic structure of the language. Accordingly, the primary recognition process uses both visual features in preperceptual storage and knowledge of orthographic structure in long-term memory during the synthesis of letter strings.

The primary recognition process operates on a number of letters simultaneously (in parallel). The visual features detected at each spatial location of the letter string define a set of possible letters for that position. The primary recognition process chooses from this set of candidates the letter alternative that has the best correspondence to the detected visual features. However, the selection of a letter can be facilitated by the reader's knowledge of orthographic structure. We assume that orthographic structure is utilized in the following manner. Upon presentation of a letter string, the primary recognition process begins integrating and synthesizing featural information passed on by feature detection to preperceptual visual storage. Featural information is resolved at different rates, and there is some evidence that gross features are available before the more detailed features (Massaro & Schmuller, 1975). As a result, the primary recognition process is faced with a succession of partial information states. These partial information states are supplemented with knowledge about orthographic structure. Assume, for example, an initial *th* has been perceived in a letter string, and the features available for the next letter eliminate all alternatives except *c* and *e*. The primary recognition process would synthesize *e* without waiting for further visual information, since initial *the* is acceptable, whereas initial *thc* is not.

The primary recognition process transmits letter information to synthesized visual memory. Figure 1 shows how the *secondary recognition* process transforms this synthesized visual percept into a meaningful form in generated abstract memory. We assume that secondary recognition attempts to translate the letter string into a word. The secondary recognition process makes this transformation by finding the best match between the synthesized letter string and a word in long-term memory. Each word in long-term memory contains both perceptual and conceptual codes. The word that is recognized is the one whose perceptual code gives the best match and whose conceptual code is more appropriate in a particular context. Analogous to primary recognition, knowledge of orthographic structure also can contribute to secondary recognition; word recognition can occur without complete recognition of all of the component letters.

Given the letters *bea* and the viable alternatives *b* and *d* in final position, only *d* makes a word, and therefore word identification (lexical access) can be achieved (Massaro, 1977).

II. Orthographic Structure and Recognition

Our goals in the present series of experiments are to provide a better understanding of primary and secondary recognition and to determine which aspects of orthographic structure the reader knows and uses. To assess how readers utilize knowledge about the structure of written language, it is necessary to state various descriptions of this structure and then to determine how well these descriptions capture reading performance. Venezky and Massaro (1979), Massaro, Venezky, and Taylor (1979), and Massaro *et al*. (1980b) have defined two broad categories of orthographic structure: *statistical redundancy* and *rule-governed regularity*. The first category includes all descriptions derived solely from the frequency of occurrence of letters and letter sequences in written texts. The second category includes all descriptions derived from the phonological constraints in English and from scribal conventions for the sequence of letters in words. Although a change in one category would not necessarily affect the other, the two categories overlap to some degree. Even with some overlap in these categories, it is of interest, first, to learn whether one of these categories reflects the manner in which readers store knowledge of orthographic structure and, second, to determine precisely which specific description within that category has the most psychological reality.

Massaro *et al*. (1979, 1980b) contrasted a specific statistical-redundancy description with a specific rule-governed description by comparing letter strings that varied orthogonally with respect to these descriptions. The statistical-redundancy measure was the summed token position-sensitive single-letter frequency. The rule-governed regularity measure was a preliminary set of rules similar to those presented in Table I of this article. Nonword letter strings were selected that represented the four combinations formed by a factorial arrangement of high or low frequency and regular or irregular. In a series of experiments utilizing a target-search task, subjects were asked to indicate whether or not a target letter was present in each of these letter strings. This task is assumed to measure the visual recognition of the letters in the letter string. Both accuracy and reaction-time measures indicated psychological reality for both the single-letter frequency and our regularity descriptions of orthographic structure.

Massaro *et al.* (1980b) formalized the language processing model to provide a quantitative description of the facilitative effect of orthographic structure on task accuracy. The basic assumption of the model is that knowledge of orthographic structure contributes an independent source of information about the letter string. By an independent source of information, we mean that knowledge of orthographic structure does not modify or direct the feature detection process. Rather, information about visual features and orthographic structure accumulates from sources that do not interact. Since information about structure supplements featural information, fewer visual features are necessary to resolve well-structured than poorly structured strings. The model was applied to the target-search task by formalizing a decision algorithm assumed to be used by the subject when faced with partial information. The model provided a good quantitative description of the accuracy results. The parameters of the model were psychologically meaningful, and the parameter values corresponding to the number of letters seen in the test string provided a quantitative measure of the contribution of orthographic structure. According to the model, readers were able to recognize two additional letters in brief presentations of well-structured strings as compared with poorly structured strings. This is a substantial effect, considering that two letters represent one-third of the six-letter test string. These results indicate that we had developed good initial approximations of both a description of orthographic structure and the means by which structure and visual features combine during word recognition. This bolstered our hope that a precise description of orthographic structure can eventually be determined and that a thorough understanding of the word recognition processes in reading can eventually be obtained.

The factorial design of the Massaro *et al.* (1980b) experiments contrasted just one measure of rule-governed regularity with one measure of statistical redundancy. Therefore, a large number of post hoc correlational analyses was conducted to evaluate a wide range of measures of orthographic structure. This was a first step toward refining our initial measures of orthographic structure. Through these correlations, it was possible to determine the refinements needed to reach our goal of a psychologically real description of orthographic structure. The dependent measure was the performance on each of 200 test items. Position-sensitive summed log bigram frequency provided the best statistical-redundancy description of performance on the individual items. Furthermore, an improved rule-based regularity measure also provided a very good description. However, the description given by the regularity measure correlated very highly with that given by the best frequency-based measure. For this reason, it was not possible in these experiments to decide whether either regularity or statistical redundancy is sufficient to account for the reader's utilization of orthographic structure.

Our goal in the present experiments is to refine our measures of structure in a further attempt to contrast a more powerful statistical-redundancy measure with an improved rule-governed regularity measure. Although statistical redundancy and regularity are highly correlated, a design involving orthogonal contrasts might be sufficient to distinguish between them. We follow this logic in the present studies by factorially contrasting bigram frequency and regularity measures in a target-search task. Words of high and low word frequency are also included as test items in order to assess the role of lexical status and word frequency. As with the previous experiments (Massaro *et al.*, 1980b), it again will be necessary to examine post hoc correlations to determine whether some other measure might provide a better description. By refining and repeatedly testing measures of structure, we hope to determine those properties that best reflect the reader's knowledge of orthographic structure.

A. Experiments 1 and 2: Bigram Frequency vs Regularity

1. *Method*

a. Subjects. Nine subjects were used in the first experiment and eleven were used in the second. All were Introductory Psychology student volunteers who received credit toward their course grade for participating. Additionally, they were all native English speakers, right-handed, had normal or corrected to normal vision, and had not participated in any of the other experiments.

b. Stimuli and Apparatus. A sample of high-frequency words was obtained from a list of all six-letter words from Kucera and Francis (1967), subject to the constraints that the words had a frequency greater than or equal to 50, were not proper nouns, and did not have repeated letters. A similar list of words with a frequency of exactly three was used to obtain low-frequency words. For each word in these two lists, all possible 720 anagrams were generated and each of their summed-positional bigram frequencies was calculated. The bigram frequencies were based on counts given by Massaro *et al.* (1980b), which were derived from the Kucera and Francis (1967) word list. Forty high-frequency and 40 low-frequency words were selected along with four anagrams of each word. The anagrams were selected so that they formed a factorial arrangement of high and low summed-positional bigram frequency and of being orthographically regular and irregular. Orthographic regularity was manipulated in

TABLE I

THE RULES FOR CHOOSING REGULAR AND IRREGULAR LETTER STRINGS

Letter strings were regarded as regular if they were phonologically legal and contained common vowel and consonant spellings. A letter string was regarded as orthographically irregular if it contained at least one of the following spellings:

a. Phonologically illegal initial or final cluster (e.g., *rlh*ued or eigo*pn*)
b. Orthographically illegal spelling for an initial final consonant or consonant cluster (e.g., *xe*oich or tmor*eh*)
c. An illegal vowel spelling (e.g., c*aei*nm)
d. A phonologically illegal medial cluster (e.g., i*lrm*ed)

the same manner as in previous experiments (Massaro *et al.*, 1979, 1980b). The rules for choosing regular and irregular strings are given in Table I. Some examples of the words and their respective anagrams are presented in Fig. 2. *Number* and *person* have high word frequencies, and *hurdle* and *pigeon* have low frequencies. The letter string *rumben* is a regular-high anagram of the word *number*, and *helrud* is a regular-low anagram of *hurdle*. The number in each cell gives the average summed-positional bigram frequency for the items of that class. For example, the irregular-high anagrams of high-frequency words have an average count of 5738.

		Summed Positional Bigram Frequency: High	Low
Words	number person (6060) / hurdle pigeon (4468)		
Orthographic Regularity	Regular	rumben preson (5624) / hulder gopine (4861)	runemb roneps (1415) / helrud ginope (1106)
	Irregular	bemrnu npsore (5738) / rlhued eiopng (4920)	brnemu pnseor (1424) / ldeurh eigopn (1113)

Fig. 2. Example of the test words and their corresponding anagrams from Experiments 1 and 2. Within each of the five squares, the top two items correspond to high word frequency and the bottom two items correspond to low word frequency. The number in each of the ten cells represents the summed-positional linear bigram frequency.

Twenty arbitrarily chosen high-frequency words and their anagrams as well as 20 low-frequency words and their anagrams were selected as stimuli for the first experiment. The remaining 20 high- and 20 low-frequency anagrams were used with new subjects in the second experiment. The letter strings for the two experiments are presented in Massaro, Jastrzembski, and Lucas (1980a).

The visual displays were generated by a DEC LSI-11 computer under software control and presented on Tektronix Monitor 604 oscilloscopes (G. A. Taylor, Klitzke, & Massaro, 1978a, 1978b). These monitors employ a P31 phosphor that decays to .1% of stimulated luminance within 32 msec of stimulus offset. The alphabet consisted of lower-case letters without serifs resembling the type font Univers 55. For an observer seated comfortably at an experimental station, the six-letter displays subtended about 1.9 degrees of visual angle horizontally, and the distance from the top of an ascender to the bottom of a descender was about .4 degree. Up to four subjects could be tested in parallel in separate rooms.

c. Procedure. A trial (see Fig. 3) began with the presentation of a 250-msec fixation point. The fixation point was replaced by a test letter string, i.e., a word or an anagram, for a duration of 10–39 msec. The duration on a particular trial for each subject was determined by his or her accuracy. The duration was adjusted every 20 trials by a modified version of the PEST algorithm (M. M. Taylor & Creelman, 1967) in order to keep the subject's average accuracy at about 75%. A masking stimulus followed the onset of the test string after a 70-msec interval. Therefore, the blank interval between the test stimulus and the masking stimulus was (70 − t) msec, where t was the duration of the target string. The masking stimulus was composed of six nonsense letters. Each nonsense letter changed from trial to trial and was composed of a montage of randomly selected features of the test letters. The feature density of a nonsense letter was equal to that of the letter g. The size of the nonsense letters was equivalent to that of the test string. The duration of the mask was adjusted along with the duration of the test string. The mask remained on the screen for (40 − t) msec, giving a range of durations of 1–30 msec. The

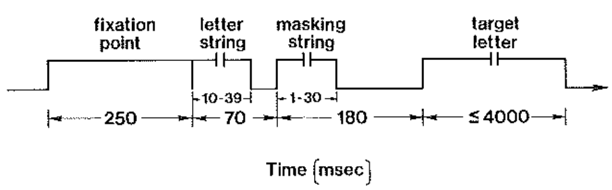

Fig. 3. A schematic representation of the perceptual recognition task used in Experiments 1-6.

mask was followed by another blank interval and then the target letter. The second blank interval lasted 180 msec minus the duration of the mask. Therefore, the interval between the onset of the test letter string and the target letter was always 250 msec. The target letter remained on the screen until all subjects responded or for a maximum of 4 sec. Finally, the interval between trials was 500 msec.

Subjects were instructed to indicate whether the target letter was present in the test string and to be as accurate as possible. The experiment consisted of a session of 100 practice trials with a practice list that was comparable to the experimental list, and two sessions of 400 experimental trials each. Within each session, each item was tested once as a target string and once as a catch string. On target trials, the target letter was selected randomly with replacement from the six letters in the test string. For catch trials, a target was selected randomly from the set of 26 letters weighted by their probability of occurrence in the stimulus set. If the selected letter was present in the test string, additional drawings with replacement were made until an appropriate target letter was selected. Some letters did not occur in any of the test strings and therefore were never tested. A short rest break intervened between the two experimental sessions. The total time for the three sessions and the rest break was about 75 min. Both experiments were conducted in exactly the same manner except that different subjects and different items were used in each.

2. *Results*

a. Analyses of Variance. Two analyses of variance were performed on the percentage accuracy scores. In the first analysis, word frequency, type of test letter string, target or catch trial, and subjects were factors. In the second analysis, the word data were eliminated, and regularity and bigram frequency were factors in the design. Figure 4 shows the average percentage correct on target and catch trials as a function of letter-string type in Experiment 1. There were large differences among the various types of letter strings, $F(4,32) = 130.7$, $p < .001$. Regular items resulted in a 9.3% accuracy advantage over irregular items $F(1,8) = 74.7$, $p < .001$; and items of high summed-positional bigram frequency had a 2.5% advantage over items of low summed-positional bigram frequency, $F(1,8) = 11.4$, $p < .01$. The advantage of high bigram frequency was limited to regular items, $F(1,8) = 10.0$, $p < .05$. The difference in accuracy between words and the regular-high anagrams was 12.0%, $F(1, 32) = 23.3$, $p < .001$. There was no significant difference in accuracy between target (72.4%) and catch (77.2%) trials, $F < 1$, and this variable did not interact with letter-string type, $F < 1$.

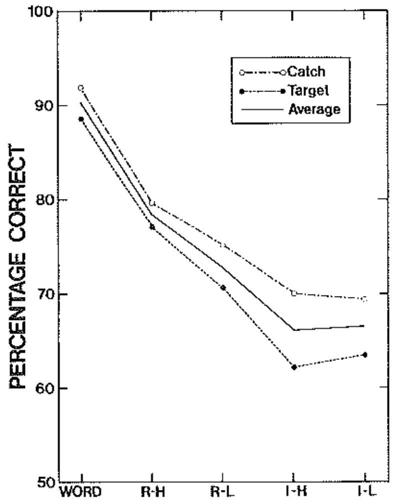

Fig. 4. Percentage correct as a function of display type for target and catch trials in Experiment 1.

Figure 5 gives the average percentage correct for the high and low word frequency words and their anagrams as a function of letter-string type. There was an overall 2.7% advantage for the low-frequency words and their anagrams, $F(1,8) = 9.86$, $p < .015$, and word frequency also interacted with letter-string type, $F(4,32) = 6.18$, $p < .001$. The overall effect of letter-string type was 28.3% for the high-frequency words and their anagrams and 19.5% for the low-frequency words and their anagrams. This difference reflected the fact that high-frequency words were more accurate than low-frequency words, but that the reverse was the case for the four types of anagrams. Word frequency did not interact with target vs catch trials, nor was there a three-way interaction with these variables and letter-string type (Fs < 1).

Figure 6 gives performance for target and catch trials as a function of letter-string type in the second experiment using new items and new subjects. There were large differences among letter-string types, $F(4,40) = 92.76$, $p < .001$. Regular items resulted in 8.7% greater accuracy than did irregular items, $F(1,10) = 49.1$, $p < .001$. Items of high summed-positional bigram frequency resulted in 3.3% greater accuracy than did items of low summed-positional bigram frequency, $F(1,10) = 12.2$, $p < .05$, but the advantage occurred only for regular items, $F(1,10) = 8.2$, $p < .025$. The difference between words and regular-high anagrams was 13.1%, $F(1,40) = 18.0$, $p < .001$). There was no significant difference in accuracy between target (74.1%) and catch (78.4%) trials, $F < 1$, and this variable did not interact with letter-string type, $F < 1$.

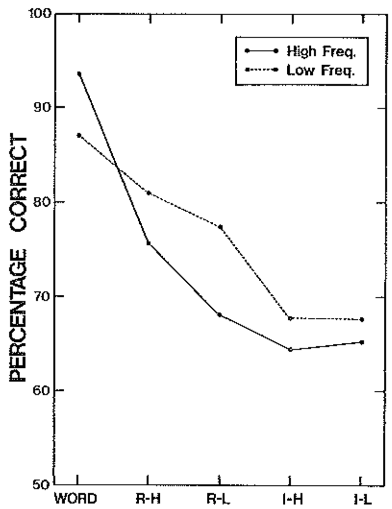

Fig. 5. Percentage correct as a function of display type for items corresponding to high and low word frequency in Experiment 1.

Figure 7 gives average percentage correct for the high and low word frequency words and their anagrams as a function of letter-string type. There was a 2.5% advantage for items of the high-frequency words, $F(1,10) = 4.87$, $p < .052$, but word frequency did not enter into any interactions. The overall effect of letter-string type was 24.3% for high-frequency items and 25.8% for low-frequency items. The interaction of

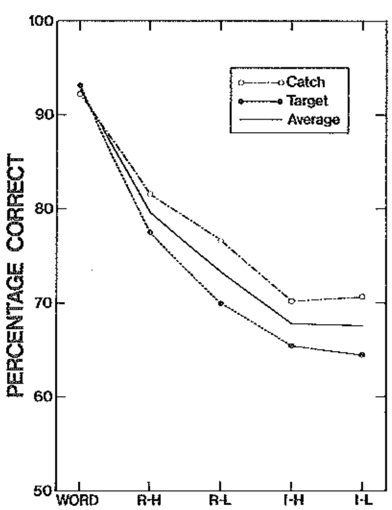

Fig. 6. Percentage correct as a function of display type for target and catch trials in Experiment 2.

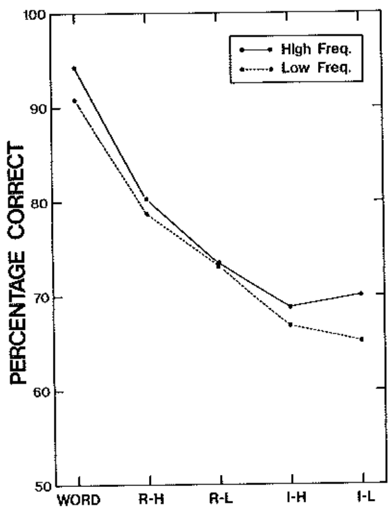

Fig. 7. Percentage correct as a function of display type for items corresponding to high and low word frequency in Experiment 2.

word frequency and letter-string type found in the first experiment and shown in Fig. 4 was not replicated in the second experiment, $F(4,40) = 1.15$, $p < .25$.

b. Correlations and Regressions. The factorial design is limited in terms of providing a quantitative assessment of the importance of frequency and regularity measures of orthographic structure. The present design contrasted just one frequency measure against just one regularity measure. Therefore, post hoc correlational analyses were carried out to provide an analysis of a range of descriptions of orthographic structure. The independent variables used in this analysis included a number of measures based on frequency counts for letters, *n*-grams, and words, in addition to a few quantitative measures based on orthographic rules. The dependent measure in all cases was average accuracy for each six-letter test item. The accuracy scores were obtained by averaging across subjects and across target and catch trials. Each of the two experiments used 40 words, 20 each of high and low word frequency, and four corresponding anagrams for a total of 200 stimulus items per experiment. Each subject had been presented with each item twice as a target trial and twice as a catch trial. Accordingly, the accuracy score for each item in the first experiment was based on 36 observations (4 replications × 9 subjects), and the accuracy score for the second experiment was based on 44 observations (4 replications × 11 subjects).

c. Frequency Measures. The source of the frequency measures is based on a word corpus compiled by Kucera and Francis (1967). This corpus consisted of 500 samples of approximately 2000 words each selected from 15 categories. Massaro *et al.* (1980b) used these words to derive the frequencies of occurrence of single letters, bigrams, and trigrams. Tables were prepared by counting the occurrence of each *n*-gram at the position it occurred in words of a given length. The counts were token counts based upon the total number of occurrences of the words containing the *n*-gram. A position-insensitive count (but still word length dependent) was also obtained for each *n*-gram by summing across the position-dependent counts. The single-letter tables and bigram tables for word lengths 3 through 7 are presented in Massaro *et al.* (1980b).

Type counts are based on the number of word types that contain a given *n*-gram; these counts may also be relevant descriptors of frequency-based measures of orthographic structure (Solso & King, 1976). However, Massaro *et al.* (1980b) found that the correlations between comparable type measures and token measures were very high. Measures based on single letters, bigrams, and trigrams, both position sensitive and position insensitive, correlated between .84 and .99. With such high correlations no meaningful discrimination between type and token measures can be made unless test items are selected with this contrast in mind. For this reason we shall discuss only measures based on the token counts derived by Massaro *et al.* (1980b).

The present analysis will be restricted to position-sensitive counts. Massaro *et al.* (1980b) found that position-sensitive counts give consistently better descriptions of performance than do position-insensitive counts. For single-letter frequency, for example, the correlation with average accuracy was only .2 for position-insensitive counts but .62 for position-sensitive counts. The advantage of position sensitivity was attenuated, however, as the length of the *n*-gram increased.

While the effects of frequency seem to be psychologically real, it is not necessary that the mental representations of frequency directly reflect the frequency of objective counts. One alternative scale that has been successful in other research is a logarithmic (base 10) scale (Solomon & Postman, 1952; Taylor, 1977; Travers & Olivier, 1978). Furthermore, a logarithmic representation is consistent with recent studies of number representation (Shepard & Podgorny, 1978) and with many other psychological scales. Therefore, we computed all our frequency measures on the basis of both linear frequencies and log frequencies. Since counts were sometimes zero, the log of zero was defined as zero. Therefore, the two sets of measures being correlated were sums of position-dependent single letters, bigrams, and trigrams derived from either linear-frequency or log-frequency tables.

d. Regularity Measures. To provide a quantitative measure of the regularity of each of the 400 stimulus items, a simple count of the number of orthographic irregularities for each item was computed, based on the rules developed by Massaro *et al.* (1980b). The rules are given in Table II. This measure of regularity provided a reasonable description of performance in the Massaro *et al.* (1980b) studies. We shall refer to this measure as Regularity(1). One critical feature of the rules for Regularity(1) is that letter strings are treated as monosyllabic and many legal and occurring medial consonant clusters are treated as irregular. For example, the word *person* would be considered to have an irregularity, since according to rule 2 the medial consonant cluster *rs* would not be legal in initial position. However, the consonant cluster *rs* is regular in medial position when considered as part of a two-syllable word. Therefore, a second quantitative measure of regularity was derived that removed the constraint that the letter string must be considered as a monosyllabic string. This measure is referred to as Regularity(2).

The rules for Regularity(2) were identical to those for the first measure except that the application of the rules and the counting of the violations were carried out in order to minimize the number of violations for any given letter string. When possible, a syllable boundary was assumed in order to avoid a given violation. As an example, the medial consonant cluster *md* in the string *limder* would be an illegal consonant cluster in the

TABLE II

THE RULES FOR AN IRREGULARITY COUNT[a]

1. Segment string into vowel and consonant substrings. Treat final *-le* as if it were *-el*. Treat *h* between vowels as a (legal) consonant.
2. For each consonant string, determine minimal number of vowels that must be inserted to make the string pronounceable. Initial consonant clusters must be legal in initial position. Final consonant clusters must be legal in final position, including those followed by final *e*. Medial consonant clusters must be legal in initial position.
3. Rate each resulting consonant substring for position-sensitive scribal regularity (count one for each irregular substring).
4. For each vowel substring, determine minimal number of consonants that must be inserted to create scribally regular sequences. Mark as irregular illegal initial and final vowel substrings.
5. Count number of inserted vowels and consonants plus number of scribally irregular consonant and vowel substrings. This yields an irregularity index.
6. The vowel string *ao*, *ae*, *oe*, and *ye* (among more obvious cases) would be illegal vowel strings. *y* would be illegal as a vowel in initial position, and *i*, *u*, *a*, *oa*, and *o* would be illegal in final position. *ue* is legal as is y as a single, noninitial vowel.
7. *h* is not allowed in final position unless preceded by *c*, *g*, or *s*.
8. *y* and *w* between vowels are to be counted as consonants.

[a] After Massaro, Taylor, Venezky, Jastrzembski, and Lucas (1980b).

same syllable because of the phonological rule governing the place of articulation of nasals followed by stops in a single syllable. The nasal and the following stop must share place of articulation; therefore *mb* and *nd* are possible, but not *md* or *nb*. A syllable boundary between *m* and *d* in *limder* is possible, however, resulting in a perfectly legal two-syllable string with no violations. Similarly, in the string *nurdgi* the medial consonant cluster *rdg* is legal with a syllable boundary between *d* and *g*. The only violation is *i* in final position.

e. Frequency vs Regularity. The correlations of several measures with average accuracy are presented in Table III. The correlation needed for statistical significance at $p = .01$ with 198 degrees of freedom is .18. Of central interest is the relative ability of bigram frequency and regularity measures of orthographic structure to predict performance. Two dummy variables were created to contrast these two measures while equating for the range and levels of each measure. The dummy regularity variable assigned a 1 to words and regular nonwords, and a 0 to irregular nonwords. The dummy frequency variable assigned a 1 to words and high bigram frequency nonwords, and a 0 to low bigram frequency nonwords.

TABLE III

CORRELATIONS OF SEVERAL PREDICTOR VARIABLES WITH OVERALL ACCURACY PERFORMANCE IN EXPERIMENTS 1 AND 2

	Experiment 1	Experiment 2
Dummy regularity	.49	.51
Dummy frequency	.26	.37
Single letter		
Linear	.32	.28
Log	.36	.49
Bigram		
Linear	.35	.29
Log	.54	.63
Trigram		
Linear	.47	.46
Log	.60	.64
Word frequency		
Linear	.45	.51
Log	.55	.59
Dummy	.60	.68
Regularity count(1)	.52	.51
Regularity count(2)	.54	.55

In both experiments the regularity variables correlated much higher (.49, .51) with performance than did the frequency variable (.26, .37).

It is not possible to choose between regularity and frequency measures of orthographic structure. Although the regularity counts give higher correlations than do linear frequency counts and log single letter counts, the log bigram and log trigram counts give the highest correlations. Both frequency and regularity measures account for a significant portion of the variance in performance. Regularity and frequency measures are positively correlated with each other. As an example, log trigram frequency and Regularity(2) correlate .47 and .46 for the items in Experiments 1 and 2, respectively. A multiple regression was carried out treating the summed frequency counts and the irregularity counts as independent variables. The best combination of predictors was log trigram frequency and Regularity(2), which accounted for 45% of the variance in Experiment 1 and 49% in Experiment 2.

f. Word Frequency. Linear word frequency correlated .45 and .51 and log word frequency correlated .55 and .59 with performance in the two experiments. The correlations with performance on just the 40 word items were .46 and .51 for linear and log frequencies in the first experiment and .34 and .35 in the second experiment. Although it is possible that word frequency makes an independent contribution to performance, the high correlations between word frequency and sublexical orthographic structure measures preclude resolution of this issue. Log word frequency was highly correlated with both log bigram frequency (.50, .52) and log trigram frequency (.77, .74). Lexical status rather than word frequency may be the critical variable producing the large advantage for word items. A dummy word frequency variable which assigned a 1 to words and a 0 to nonwords was more highly correlated (.60, .68) with performance than was log word frequency.

B. Experiment 3: Log Bigram Frequency vs Regularity

The creation of the stimulus set was identical to that of the previous two experiments except that log bigram rather than linear bigram counts were used and the strings were controlled more exactly for regularity. Figure 8 gives examples of the five classes of items and the average log bigram frequency for each class. The complete list of letter strings is presented in Massaro *et al.* (1980a).

In the studies of Massaro *et al.* (1980b) and Experiments 1 and 2, log-frequency measures were consistently more highly correlated with performance than were linear-frequency measures. Furthermore, Massaro

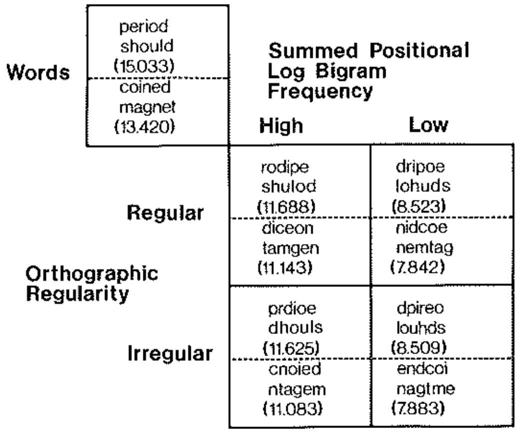

Fig. 8. Examples of the test words and their corresponding anagrams from Experiments 3 and 4. Within each of the five squares, the top two items correspond to high word frequency and the bottom two items correspond to low word frequency. The number in each of the ten cells represents the summed positional log bigram frequency.

et al. (1980b) found that the log counts were superior to a range of power-function transformations of the linear counts. This result provides additional evidence that, if frequency of occurrence is important, log frequency appears to be the best descriptor of this variable.

A count of the number of irregularities in each letter string was determined by using the rules for Regularity(3) presented in Table IV. The

TABLE IV

THE RULES FOR REGULARITY (3) FOR THE SELECTION OF THE ITEM USED IN EXPERIMENTS 3–6

1. For each string, rate for position-sensitive scribal regularity and pronounceability (count one for each violation). Treat final *-le* as if it were *-el*. Treat *h* between vowels as a (legal) consonant.
2. For each string, rate for position-sensitive scribal regularity and pronounceability (count one for each violation). Initial consonant clusters must be legal in initial position. Final consonant clusters must be legal in final position, including those followed by final *e*.
3. For each string, rate for position-sensitive scribal regularity and pronounceability (count one for each violation). The vowel strings *ao, ae, oe,* and *ye* (among more obvious cases) would be illegal vowel strings. All vowels are illegal in initial position and *y* would be illegal as a vowel in initial position. The vowels *i, u, a, oa,* and *o* would be illegal in final position. *ue* is legal as is *y* as a single noninitial vowel. *h* is not allowed in final position unless preceded by *c, g,* or *s*. *y* and *w* between vowels are to be counted as consonants.

rules for Regularity(3) were the same as those for Regularity(2), except that vowels as initial letters violated one of the rules and therefore were counted as irregularities. Given this formula, it was now possible to equate the number of irregularities for the anagrams that differed only in log bigram frequency. In our previous studies, the number of irregularities tended to correlate negatively with frequency, and some of the effect of frequency could have been due to differences in regularity. This possibility was eliminated in the present study by equating the high- and low-frequency anagrams of a given test word for the number of irregularities. Consider the test word *period* shown in Fig. 8. The regular-high and regular-low anagrams (*rodipe* and *dripoe*) do not have any irregularities. The irregular-high and irregular-low anagrams (*prdioe* and *dpireo*) have two irregularities each. This design might provide a more definitive contrast between frequency and regularity.

1. Method

a. Subjects. Nine University of Wisconsin summer school student volunteers were used as subjects and paid $9.00 for their participation. All were native English speakers, had normal or corrected vision, and had not participated in any of the other experiments.

b. Stimuli and Apparatus. Words were selected in the same manner as in the previous two experiments. The high-frequency words had a Kucera and Francis (1967) frequency of at least 50, and the low-frequency words had a frequency of 3. Because of a selection error, one low-frequency word had a frequency of 4. For each of the 80 words, four anagrams were selected so that they formed a factorial arrangement of high and low summed-positional log bigram frequency and of being orthographically regular or irregular. For each set of four anagrams, the number of irregularities were matched exactly for the regular conditions and then again for irregular conditions. Finally, an additional sample of words, 13 high and 13 low in word frequency, and their anagrams were selected as practice items.

The 80 experimental words and their anagrams were divided into two lists. List 1 contained one-half of the high word frequency items and one-half of the low word frequency items. List 2 contained the remaining items.

Stimuli were presented in the same manner and on the same equipment as in the previous experiments with only one exception. The range of durations for the test letter strings was 5–39 msec. Because of the algorithm used, decreasing the lower limit for the duration of the test string increased the maximum duration of the mask to 35 msec.

c. Procedure. The experiment was conducted in a manner similar to that of the previous experiments. The presentation of the test string, masks, and target letters was identical to that of Experiments 1 and 2. Subjects were tested on two consecutive days. At the beginning of each day, subjects began with a practice session of 260 trials. Two experimental sessions of 400 trials each followed the practice. Five of the subjects received all 200 items of List 1 on Day 1 in the first session as both target and catch trials. These subjects then received the List 2 items in the second session. On Day 2, List 2 was presented in the first session and List 1 in the second session. For the remaining four subjects, the order of the lists was reversed.

2. *Results*

a. Analyses of Variance. Figure 9 shows the average percentage correct on target and catch trials as a function of letter-string type. There were significant differences, $F(4,32) = 127.1$, $p < .001$, among the five types of letter strings. Words had a 16% advantage over the regular-high anagrams, $F(1,32) = 55.1$, $p < .001$. There was 4.0% advantage of regular strings over irregular strings, $F(1,8)$ 32.5, $p < .001$, and a 1.4% advantage of high log bigram frequency strings over low log bigram frequency strings, $F(1,8) = 2.6$, $p > .2$.

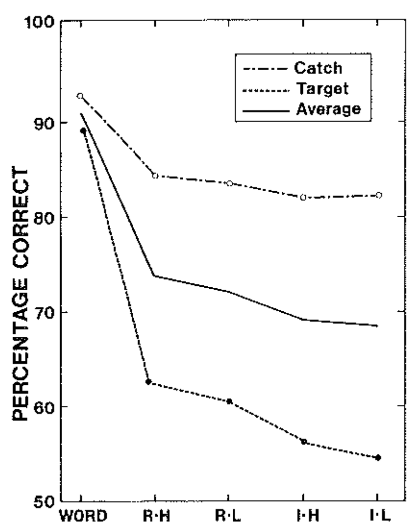

Fig. 9. Percentage correct as a function of display type for target and catch trials in Experiment 3.

One disquieting aspect of the results is the extreme asymmetry in performance on target and catch trials and the interaction of this variable with display type, $F(1,8) = 12.8$, $p < .007$, and $F(4,32) = 9.5$, $p < .001$. Subjects were extremely conservative in their willingness to indicate that a target letter was present. This result could reflect our failure to instruct the subjects specifically about the relative frequency of target trials as we did in the previous two experiments.

Figure 10 gives average percentage correct for the high and low word frequency words and their anagrams as a function of letter-string type. High-frequency words and their anagrams were recognized 3.2% more accurately than were low-frequency words and their anagrams, $F(1,8) = 69.03$, $p < .001$. The interaction between word frequency and the five types of items was not significant, showing that this difference was not unique to the word items. Therefore, some variable other than word frequency must be responsible for the difference. However, one caveat is to realize that performance may not be on an interval scale, which weakens any interpretation of the lack of interaction. One solution would be to monitor each display type independently and to adjust the stimulus values to give an average of 75% correct for each display type. If word frequency still does not interact with display type when average performance is about 75% correct at each display type, then the conclusion reached here would be reinforced.

b. Correlations and Regressions. The correlations of several variables with overall performance are presented in Table V. For all fre-

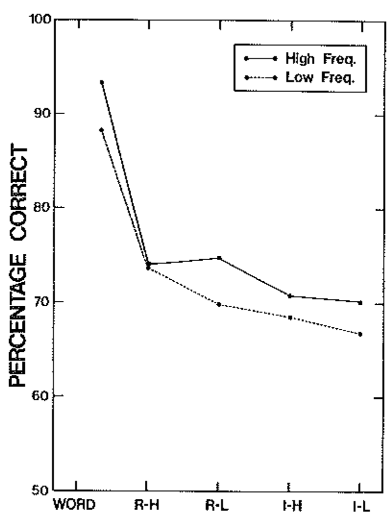

Fig. 10. Percentage correct as a function of display type for items corresponding to high and low word frequency in Experiment 3.

TABLE V

CORRELATIONS OF SEVERAL PREDICTOR VARIABLES WITH OVERALL ACCURACY IN EXPERIMENTS 3 AND 4

	Exp. 3	Exp. 4	Ave.
Single letter			
Linear	.29	.26	.31
Log	.36	.28	.36
Bigram			
Linear	.38	.33	.40
Log	.53	.43	.54
Trigram			
Linear	.44	.36	.45
Log	.59	.48	.60
Word frequency			
Linear	.44	.37	.46
Log	.60	.50	.63
Regularity count(3)	.40	.29	.39

quency measures, the log measures correlated more highly with performance than did the linear measures. Log trigram frequency predicted performance better than did the other sublexical measures. Log word frequency was correlated with accuracy (.60), but also was correlated with log bigram frequency (.61) and log trigram frequency (.80). Among just the high-frequency words the correlation with performance was −.05 and −.13, respectively, for linear and log word frequencies. (Correlations among the low-frequency words would not be meaningful, since all the items had the same Kucera and Francis frequency of occurrence.) The lack of a significant correlation between performance and word frequency within the class of words replicates previous results (Manelis, 1974) and makes it unlikely that word frequency can account for the effects of orthographic structure. Lexical status alone might be an important variable, however. The dummy variable of word or nonword gave a highly significant correlation of .60 with performance.

Summed log trigram frequency and Regularity(3) accounted for 39% of the variance.

C. EXPERIMENT 4: REPLICATION

1. *Method*

Eight new University of Wisconsin undergraduates from Introductory Psychology who met the same requirements as in the previous experiments were used as subjects. Experiment 4 was an exact replication of Experiment 3 with one exception. The instructions were modified to inform subjects that a target letter would appear in the test string on 50% of the trials. It was expected that this manipulation would attenuate the asymmetry in the number of positive and negative responses.

2. *Results*

a. Analysis of Variance. Figure 11 shows the average percentage correct for target and catch trials for the five letter-string types. The significant differences among the letter-string types, $F(4,28) = 102.7$, $p < .001$, completely replicate Experiment 3. There was a 15.0% advantage of words over regular-high anagrams, $F(1,28) = 52.4$, $p < .001$; a 2.4% advantage for regular strings, $F(1,7) = 10.2$, $p < .025$; and only a 0.7% advantage for high-frequency strings, $F(1,7) = .84$.

Although the responding asymmetry was substantially reduced, there was still a tendency for subjects to remain conservative in their willingness to indicate that a target letter was present, $F(1,7) = 6.11$, $p < .05$. The range of performance across letter-string types was 23.0% for target trials and 13.3% for catch trials.

Figure 12 presents the percentage correct for the letter-string types as a function of word frequency. There was an overall effect of 20.5% for high word frequency items and 15.8% for low word frequency items, $F(4,28) = 2.19$, $p < .10$.

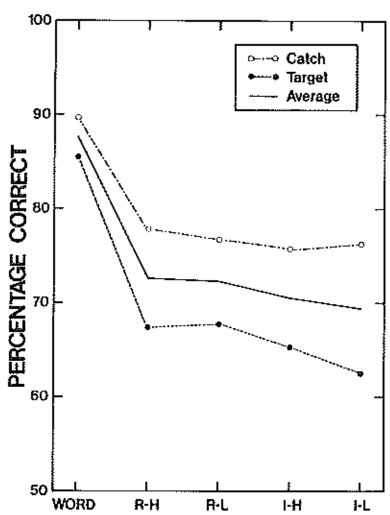

Fig. 11. Percentage correct as a function of display type for target and catch trials in Experiment 4.

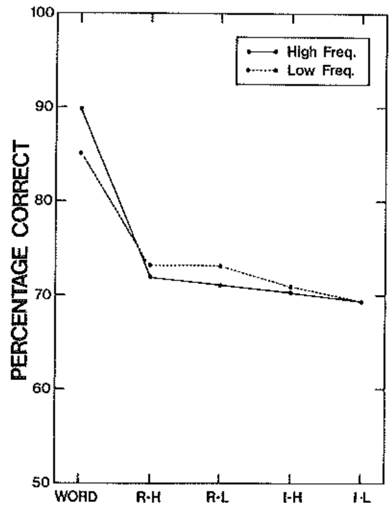

Fig. 12. Percentage correct as a function of display type for items corresponding to high and low word frequency in Experiment 4.

b. Correlations and Regressions. The correlations of several measures with overall performance in Experiment 4 are presented in Table V. As with the previous analyses, log measures predicted performance better than did linear measures. Trigram frequency was the best of the three frequency measures, but only slightly better than bigram frequency. Log word frequency was correlated .50 with overall performance.

Summed log trigram frequency and Regularity(3) accounted for 25% of the variance.

Table V also presents the correlations of the same variables with the average performance across Experiments 3 and 4. Since Experiment 4 was a replication of Experiment 3, the performance on each item was averaged across the two experiments. As might be expected from increased reliability, these correlations are significantly larger than those for either of the experiments considered separately.

D. Experiment 5: Lexical Status

In Experiments 1–4, a large advantage was found for words as compared with the best regular-high anagrams. One way to account for this effect is by the lexical status of the words. Since words are represented in the reader's lexicon, they may be retrieved on the basis of partial visual information. For example, the partial information sho-l- given presentation of *should* might lead to recognition of the word *should.* Lexical access would allow determination of the two unknown letters. On the other hand, the partial information shu-o- given presentation of *shulod* cannot access any lexical entry, and the missing letters cannot be determined. Consequently, on a word trial there is a better chance that all the component letters will be available for comparison against the target letter. In contrast, the same partial information about an anagram will not lead to recognition of all the letters in the test string. As a result, fewer letters of anagrams will be available for comparison against the target letter. This account is consistent with the model articulated in Section I; the secondary recognition process can lead to word recognition without complete recognition of all the component letters.

A second explanation of the word advantage is that words differ from even the best anagrams with respect to sublexical orthographic structure. For example, the bigram frequency of the words in Experiments 3 and 4 averaged almost three log units more than that for the regular-high anagrams (see Fig. 8). Perhaps accuracy was greater for words because words contained more frequent bigrams.

To choose between these two explanations in Experiment 5, the words of Experiments 3 and 4 were replaced with regular anagrams, which were matched with the words on log bigram frequency. If log bigram frequency was the basis of the word advantage, then a similar advantage should be observed for these regular-very high (R-VH) anagrams.

1. *Method*

Experiment 5 was conducted in the same manner as the previous experiments. Regular-very high anagrams with log bigram frequencies similar to the words of Experiments 3 and 4 were used along with all the anagrams of Experiments 3 and 4. The summed bigram frequencies for the regular-very high anagrams were 14.940 and 13.407, respectively, almost identical to those of the words (see Fig. 8). The regular-very high anagrams are listed in Massaro *et al*. (1980a). The regular-very high anagrams of *period* and *coined* were *poried* and *conied*, respectively. The experiment was identical to Experiment 4 except that the regular-very high anagrams were used in place of the word items. Seven new subjects were tested.

2. *Results*

Figure 13 shows the differences among the five types of letter strings, $F(4,24) = 3.3$, $p < .05$. There was a 0.3% advantage of regular very-high anagrams over regular-high anagrams, $F(1,6) < 1$. Regular anagrams gave a 3.6% advantage over irregular anagrams, $F(1,24) = 13.8$, $p < .005$. There was a −0.2% effect for log bigram frequency, $F(1,6) < 1$.

There was only a 3.0% advantage of catch over target trials, $F(1,6) <$

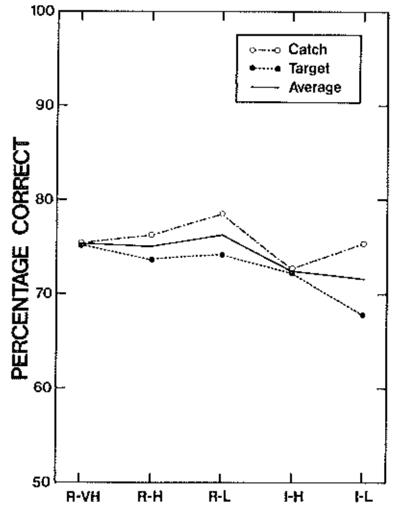

Fig. 13. Percentage correct as a function of display type for target and catch trials in Experiment 5.

1, but this variable interacted with the type of letter string, $F(4,24) = 3.3$, $p < .05$. This test reflects the presence of a 7.4% increase in accuracy from the worst to the best structured strings for target trials, but an absence of an effect of orthographic structure for catch trials (Fig. 13).

Figure 14 presents the average accuracy as a function of letter-string

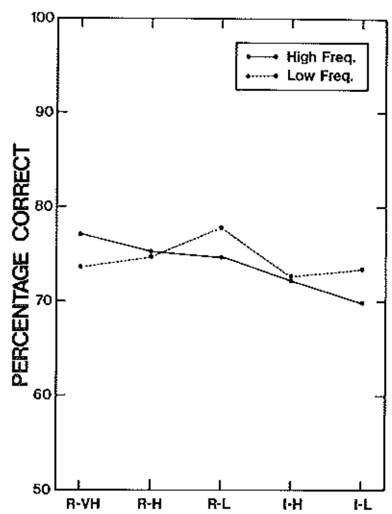

Fig. 14. Percentage correct as a function of display type for items corresponding to high and low word frequency in Experiment 5.

type according to whether the items correspond to high or low word frequency. Neither the 0.6% difference nor the interaction with display type was statistically significant.

The results of Experiment 5 support the idea that lexical status makes a significant contribution to perceptual recognition in the target search task. The contribution of lexical status in the earlier experiments cannot be attributed to sublexical orthographic structure differences in log bigram frequency. That is, the reader takes into account not only the frequency of occurrence of letter sequences and the regularity of these sequences, but also whether or not a particular sequence is represented in a word. Frequency and regularity allow well-structured anagrams to be better recognized than poorly structured anagrams; in addition, lexical status allows a perceptual advantage of words over equally well-structured anagrams.

E. Experiment 6: Frequency, Regularity, and Lexical Status

Experiments 1–5 were successful in demonstrating frequency, regularity, and lexical status as psychological measures of orthographic structure. In a final evaluation of the relative contribution of these measures, the perceptual recognition task was replicated with five display types. The display types were chosen to give a large range of regularity and bigram frequency. The comparisons among display types and the post hoc corre-

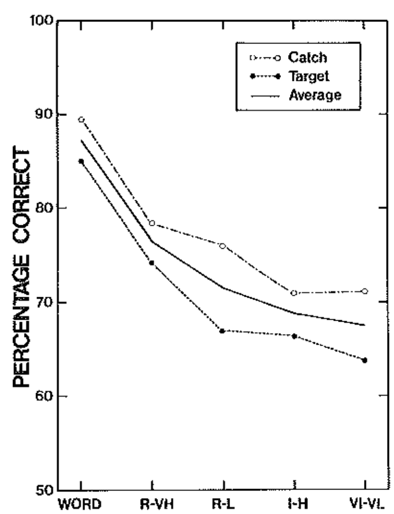

Fig. 15. Percentage correct as a function of display type for target and catch trials in Experiment 6.

lations will be used to measure the relative contributions of lexical status, regularity, and frequency in perceptual recognition.

1. *Method*

The R-VH anagrams from Experiment 5 and the words, regular-low anagrams, and irregular-high anagrams from Experiment 3 were used as items along with a new type of anagram that was both very irregular and very low in log bigram frequency. The very irregular, very low (VI-VL) anagrams mostly had three or four irregularities and average log bigram frequencies of 4.845 and 4.971 for the high and low word frequency items, respectively. The very irregular, very low anagrams are listed in Massaro *et al.* (1980a). The anagrams *pdreio* and *dcoeni* were derived from the words *period* and *coined,* respectively. The procedure of Experiment 4 was replicated exactly. Eleven new subjects from the Introductory Psychology subject pool were used.

2. *Results*

a. Analyses of Variance. As can be seen in Fig. 15, accuracy uniformly increased with better structured letter strings; $F(4,40) = 69.4$, $p < .001$. Words had a 10.8% advantage over the regular-very high anagrams, $F(1,40) = 15.9$, $p < .001$. Regular-very high anagrams gave a performance advantage of 4.9% over regular-low anagrams, $F(1,40) = 3.1$, $p < .086$; a 7.6% advantage over irregular-high anagrams, $F(1,40) = 7.9$, $p < .01$; and a 8.9% advantage over very irregular, very low anagrams, $F(1,40) = 10.7$, $p < .005$.

The 6.9% advantage of catch over target trials was not significant, $F(1,10) = 1.5$, $p > .25$, and this difference did not interact with display type, $F < 1$.

Figure 16 reveals a 2.7% difference between levels of word frequency, $F(1,10) = 32.17$, $p < .001$, but the advantage of high word frequency occurred only for the words and the regular anagrams.

b. Correlations and Regressions. The correlations of several predictor variables with overall performance are presented in Table VI. As usual, the log measures predict performance better than the linear measures do. Bigrams and trigrams were similar in predictive ability. Log word frequency correlated .42 with overall performance, but also correlated .48 with summed log bigram frequency and .75 with summed log trigram frequency.

Summed log trigram frequency and Regularity(3) accounted for 37% of the variance.

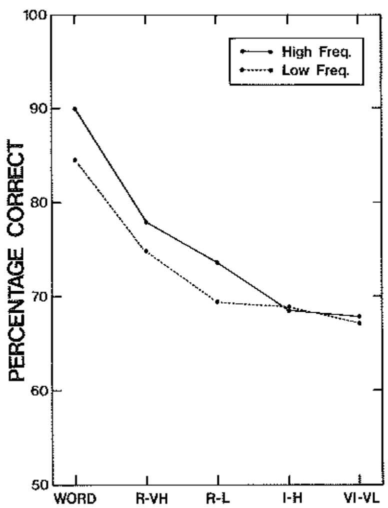

Fig. 16. Percentage correct as a function of display type for items corresponding to high and low word frequency in Experiment 6.

III. Orthographic Structure and Conscious Knowledge

The perceptual recognition task allows us to assess the degree to which readers utilize orthographic structure in visual processing of letter strings. An overt judgment task has been used to assess the degree to which this knowledge is available for a conscious report (Massaro *et al.*, 1980b; Rosinski & Wheeler, 1972). Massaro *et al.* (1980b) asked whether subjects could discriminate among the items on the basis of rule-governed regularity or on the basis of statistical redundancy. Subjects were presented pairs of letter strings and asked to choose the member of each pair that most resembled written English. The instructions emphasized either a regularity or a statistical-redundancy criterion. Subjects' judgments seemed to reflect some knowledge of both rule-governed regularity and statistical redundancy. The present experiment uses this task with our new items derived from improved frequency and regularity measures. These items should provide a more definite contrast between frequency and regularity descriptions of orthographic structure.

Experiment 7: Overt Judgments

An overt judgment task is used in the present experiment to assess the degree to which regularity and frequency are consciously available. Sub-

TABLE VI

CORRELATIONS OF SEVERAL PREDICTOR VARIABLES WITH OVERALL ACCURACY IN EXPERIMENT 6

Single letter	
Linear	.35
Log	.37
Bigram	
Linear	.40
Log	.50
Trigram	
Linear	.41
Log	.59
Word frequency	
Linear	.38
Log	.54
Regularity count(3)	.44

jects are given pairs of letter strings and asked to choose which letter string most resembles English spelling. Some subjects are instructed to base their decision on the frequency of occurrence of letter sequences in English spelling; other subjects are instructed to respond on the basis of the regularity of letter sequencing. The seven types of letter strings varying in lexical status, regularity, and log bigram frequency were paired with each other in the task. The degree to which subjects can follow instructions and discriminate among the types of items should reveal which aspect(s) of orthographic structure is (are) consciously available and capable of report.

1. *Method*

a. Subjects. Sixteen Introductory Psychology students who met the same requirements as in the first six experiments were used as subjects.

b. Stimuli and Apparatus. The 280 letter strings represented all seven categories of items used in the previous experiments. Accordingly, 40 words and their corresponding R-H, R-L, I-H, and I-L anagrams of Experiment 3, 40 R-VH anagrams of Experiment 5, and 40 VI-VL anagrams of Experiment 6 were selected. The irregular items chosen from Experiment 3 had two irregularities. Seven categories, and allowing the two letter strings of a pair to be from the same category, result in 28 unique pairs of categories. These 28 pairs were sampled randomly without replacement in each block of 28 trials. The actual items from each of the categories for each pair were randomly selected with replacement for each group of subjects. Eventually, each subject was presented with 840 pairs for judgment, resulting in a total of 30 observations for each pair. The two strings of each pair were arranged side by side on the CRT. The horizontal visual angle of each letter string was 1.9 degrees with a 2.7-degree separation between strings. Subjects selected the one of the two strings that more closely resembled an English word. Subjects indicated their choice by pressing one of two keys located beneath the string. Each trial began with a 250-msec fixation point followed by the two letter strings. The strings remained on the CRT until all the subjects responded or for a maximum of 4 sec. The 840 pairs were presented in two sessions of 420 trials each. Each session lasted about 20 min. Of the 16 subjects, 8 were given the regularity instructions and 8 were given the frequency instructions. The regularity instructions described regularity of letter sequencing, and subjects were asked to choose the more regular of the two letter strings. The frequency instructions described the frequency of letter groups, and subjects were asked to choose the more frequent of the two letter strings. The exact instructions were given in Massaro *et al.* (1980a).

2. *Results*

For each subject, the proportion of times that each of the seven categories was chosen as most like English over the other six categories was computed. These proportions were entered into an analysis of variance with instructions, category type, and subjects as factors. Figure 17 presents the percentage of choices of most like English as a function of category and instructions. There was a large decrease in choices with decreases in orthographic structure, $F(6,84) = 351$, $p < .001$. However, instructions had no influence on performance and did not interact with structure, Fs < 1. All differences between adjacent categories in Fig. 17 are statistically significant, except for the small difference between R-H and R-L items.

Table VII presents the proportion of times each of the seven classes of items was chosen over the other six classes. Some effects of instructions not apparent in the average proportions shown in Fig. 17 are seen in these results. The R-H items were picked over the R-VH items 39% of the time for regularity instructions and 25% of the time for frequency instructions. The three classes of irregular items (I-H, I-L, and VI-VL) were chosen an average of 13% of the time over the R-L items with regularity instructions and an average of 18% of the time with frequency instructions. Regular-very high items were chosen over words 17% of the time for regularity

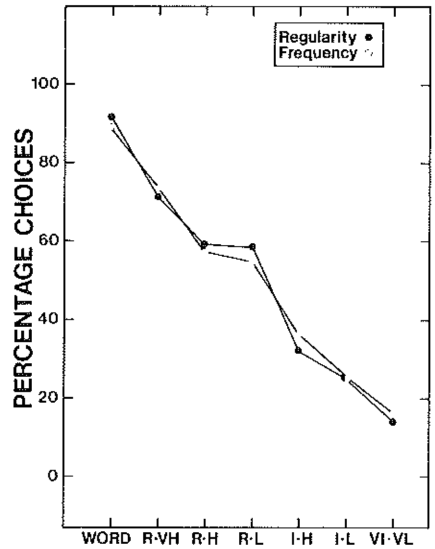

Fig. 17. Average percentage choices of each of the display types in the overt judgment task in Experiment 7.

TABLE VII

THE PROPORTION OF TIMES THE ROW ITEM WAS CHOSEN OVER THE COLUMN ITEM FOR REGULARITY AND FREQUENCY INSTRUCTIONS

Regularity instructions

	Word	R-VH	R-H	R-L	I-H	I-L	VI-VL
Word	.55						
R-VH	.17	.52					
R-H	.12	.39	.50				
R-L	.15	.32	.50	.51			
I-H	.02	.11	.24	.17	.48		
I-L	.01	.09	.15	.17	.36	.58	
VI-VL	.00	.02	.09	.06	.28	.30	.53

Frequency instructions

	Word	R-VH	R-H	R-L	I-H	I-L	VI-VL
Word	.55						
R-VH	.25	.50					
R-H	.18	.25	.57				
R-L	.11	.32	.44	.55			
I-H	.07	.14	.27	.25	.50		
I-L	.02	.09	.17	.20	.38	.52	
VI-VL	.03	.05	.09	.09	.24	.35	.49

instructions and 25% of the time for frequency instructions. These results are consistent with the idea that a dimension of orthographic structure carries somewhat more weight when that dimension is stressed in the instructions in the overt judgment task.

An analysis of variance also was conducted on the reaction times of the choice responses. Response type (selected vs. unselected) was included as a factor to assess the differences in reaction times between choosing a given category as most like English relative to the average reaction time for choosing the other six categories. Figure 18 presents the reaction times as a function of instructions, response type, and category. Overall reaction times were 232 msec longer for frequency than for regularity instructions, $F(1,14) = 6.9$, $p < .025$. Reaction times increased with decreasing orthographic structure, but only for the selected response type, $F(6,84) = 12.7$ and 21.0, $ps < .001$.

A comparison between the perception task and overt judgment task shows that the latter is a much more sensitive measure of the reader's knowledge of orthographic structure. Subjects are able to discriminate among certain classes of items in the overt judgment task that are responded to equivalently in the perceptual accuracy task. For example, R-VH and R-H were differentiated in the overt judgment task but were responded to equivalently in the perceptual accuracy task of Experiment 5. The same was true of the I-H vs. I-L and the I-H vs. VI-VL contrasts. Other results were exactly parallel in the two tasks: Words have an advan-

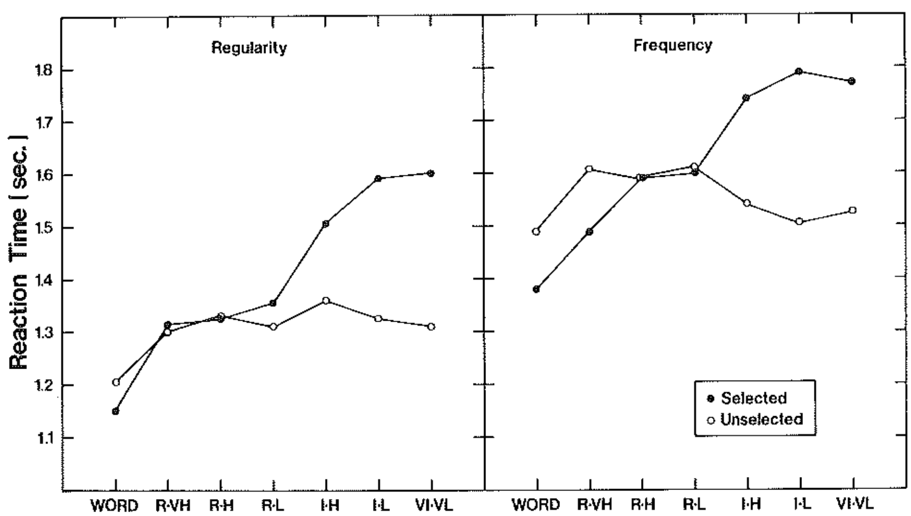

Fig. 18. Average reaction times for choosing each of the display types (selected) and for choosing the other six categories (unselected) in Experiment 7.

tage over regular items, and regular items have an advantage over irregular items.

In summary, the results from the overt judgment task paralleled the results from the target-search task. Evidently, readers not only use their knowledge of orthographic structure during the word recognition process, but also are aware of this knowledge and can use it in tasks requiring decisions after the word recognition processes have been completed. As suggested by the model, orthographic structure appears to exert an influence on several stages of language processing (Massaro, 1980).

IV. General Discussion

The orthogonal contrasts of lexical status, word frequency, position-sensitive frequency, and regularity provided evidence for the following conclusions. Lexical status produced a perceptual advantage of words over equally well-structured anagrams. Word frequency added very little, if anything, beyond that accounted for by lexical status. In the factorial design, regular anagrams were recognized significantly better than irregular anagrams, whereas log bigram frequency had no influence when regularity is controlled. However, the post hoc correlations of these measures of orthographic structure with perceptual recognition of each of the 400 test items revealed that log trigram frequency correlated very highly with performance on the individual letter strings. In this regard, frequency measures allow a fine-grained description of orthographic structure that provides a good index of performance on individual letter strings. The binary classification of lexical status and the small range of the number of irregularities limit the usefulness of these measures as descriptions of a relatively continuous variation in orthographic structure. Some frequency weighting of a regularity description might lead to an improved measure of structure. Until such a description is developed, however, it appears necessary to include lexical status, frequency, and regularity to account for those components of orthographic structure that are psychologically real. The results of the present studies also are relevant to previous studies of orthographic structure.

A. Related Research

The utilization of orthographic structure in reading was first studied by Miller, Bruner, and Postman (1954), who had subjects reproduce letter sequences presented tachistoscopically. The strings were eight letters and corresponded to different approximations to English based on Shannon's

(1948, 1951) algorithms. Miller *et al.* found that performance improved with better approximations to English. By correcting for the relative information per letter in the strings, the amount of information transmitted was shown to be equal for the various approximations. This result is consistent with more recent empirical and theoretical work demonstrating that orthographic structure provides an independent source of information to the reader (Massaro, 1979a, 1979b; Massaro *et al.*, 1980b). In fact, the approximation-to-English algorithms may be viewed as early descriptions of orthographic structure. Accordingly, the more recent studies replicate and extend the Miller *et al.* results. The major advances in the recent studies are the more precise descriptions of structure (see Massaro *et al.*, 1980b, Chap. 3) and the quantitative modeling of the processes by which visual information combines with structure during word recognition (Massaro, 1979a, 1979b).

Related research by Gibson and her colleagues evaluated the role of word length and pronounceability in a full report of letter strings by both hearing and deaf readers (Gibson, Pick, Osser, & Hammond, 1962; Gibson, Shurcliff, & Yonas, 1970). They found that the number of errors increased with increases in word length and decreased with increases in pronounceability. In the post hoc regression analyses, word length accounted for 72% of the variance and pronounceability accounted for another 15%. Position-sensitive and word-length-specific bigram and trigram frequencies were significantly poorer predictors of performance. However, these counts cannot be used in any straightforward manner for items of various letter lengths. Words of different lengths do not occur with equal frequency, and the less frequent word lengths will naturally have bigrams and trigrams with smaller counts. Therefore, this comparison cannot be considered an adequate test between pronounceability and frequency measures of orthographic structure. The finding that deaf and hearing readers were influenced similarly by pronounceability argues that orthographic structure rather than pronounceability is the important structural variable.

Manelis (1974) found an advantage of four-letter words over pseudowords in tachistoschopic recognition, but failed to find a significant correlation between recognition and summed linear bigram and trigram frequencies, as measured by Mayzner and Tresselt (1965) and Mayzner, Tresselt, and Wolin, (1965). In a more recent study, McClelland and Johnston (1977) independently varied position-sensitive bigram frequency and lexical identity in four-letter strings in a Reicher–Wheeler forced-choice task (Reicher, 1969; Wheeler, 1970). In addition, a full report of the four letters either preceded or followed the forced-choice response. The forced-choice responses revealed no effect of either bigram frequency or an advantage of words over orthographically regular pseudowords. Forced-choice responses also showed a 13% advantage of words and pseudowords over single letters, replicating the word-letter difference of Reicher (1969). The full report score replicated the absence of a bigram frequency effect found for the forced-choice task, but showed an advantage of words over pseudowords. Also, the full report score revealed a large word-frequency effect. In post hoc analyses, McClelland and Johnston report that bigram frequency did not correlate with perceptual accuracy, whereas single-letter position frequency was highly correlated with accuracy.

Using a different task, Henderson and Chard (1980) presented items either high or low in both position-sensitive single-letter and bigram frequencies in a lexical decision task. Their results indicate that second and fourth graders were faster in rejecting low-frequency than high-frequency six-letter nonwords. In a related study, Bouwhuis (1979) found that single-letter positional frequency correlated with lexical decisions for three-letter items in Dutch. Reaction times to words decreased whereas reaction times to pseudowords increased with increases in single-letter frequency. Similarly, subjects tended to respond "word" more often to both words and pseudowords if the items were of high single-letter frequency. In contrast, these correlations were considerably diminished when bigram positional frequency was used as the predictor variable. Bouwhuis' results, when compared with those of Henderson and Chard (1980), imply that the power of the bigram frequency measure with our six-letter items may not generalize completely to smaller letter-string lengths. That is, bigram frequency appears to have more predictive power than single-letter frequency when the items are six letters in length, but the reverse is indicated when the items are three or four letters in length. However, such a conclusion is tenuous for two reasons. First, single-letter and bigrams measures are highly correlated even for small letter-string lengths. Second, experiments demonstrating the predictive power of single-letter frequencies have used linear rather than log counts. Since log counts are uniformly better predictors of the data, it will first have to be shown that log counts do not change the relative power of the two frequency measures. Of the several studies (Bouwhuis, 1979; McClelland & Johnston, 1977) investigating the relative contributions of single-letter and bigram frequencies, only the present studies directly compare linear and log single-letter and bigram counts. The present studies found that log counts are consistently better than linear counts and that bigram counts are better than single-letter counts.

In the first of two experiments investigating other structural variables, Spoehr (1978) showed that report accuracy in the Reicher–Wheeler task was lower for five-letter, one-syllable strings made up of five phonemes than for those made up of four phonemes. Performance for words such as *thump* and pseudowords such as *sherk* averaged 76% correct, whereas performance was 4% worse for words such as *spank* and pseudowords such as *crost*. Average accuracy on words was 7% greater than on pseudowords. In the second experiment, two-syllable words were recognized 13% more poorly than were one-syllable words when phoneme length was equated. Although Spoehr (1978) showed that position-sensitive bigram frequency of the letters could not account for the observed differences, log counts might have been more appropriate. Furthermore, since our counts were derived from the considerably larger Kucera and Francis (1967) corpus, they are likely more reliable than the Mayzner and Tresselt (1965) counts that Spoehr employed. Accordingly, Spoehr's results are not necessarily inconsistent with the present results.

Although recent experiments have failed to find significant effects of position-sensitive bigram frequency in the perceptual recognition of letter strings, these studies all used linear rather than log counts. We have found much larger effects with log than with linear counts. Linear and log counts correlate .84 and .66 for single-letter and bigram position-sensitive counts for four-letter words, .86 and .76 for five-letter words and .85 and .76 for six-letter words in the Kucera–Francis corpus. Therefore, there is sufficient room for improvement of log over linear counts in accounting for perceptual recognition. It is necessary to evaluate log as well as linear counts to provide a sufficient test of frequency measures of perceptual recognition.

B. Summary

The present research assessed the role of orthographic structure in the perceptual recognition and the judgment of letter strings. Lexical status, word frequency, bigram frequency, log bigram frequency, and regularity of letter sequencing were varied across a series of seven experiments. Six-letter words and their anagrams were used as test stimuli in a target-search task. Words were recognized better than their corresponding equally well-structured anagrams, but word frequency had small and inconsistent effects. Orthographically regular anagrams were recognized better than irregular anagrams, whereas log bigram frequency did not have an effect. In contrast, post hoc correlations revealed that log trigram frequency did correlate significantly with individual item performance. In a final experiment, subjects judged which of a pair of letter strings most resembled English in terms of either the frequency or the regularity of letter sequences. The results revealed an influence of essentially the same dimensions of orthographic structure as was revealed by the perceptual recognition task. The results provide evidence for lexical status, regularity of letter sequencing, and frequency of letter sequencing as important dimensions in the psychologically real description of orthographic structure.

REFERENCES

Bouwhuis, D. G. *Visual recognition in words*. Unpublished doctoral dissertation, Katholieke Universiteit, Nijmegen, Netherlands, 1979.

Gibson, E. J., Pick, A., Osser, H., & Hammond, M. The role of grapheme-phoneme correspondence in the perception of words. *American Journal of Psychology*, 1962, **75**, 554–570.

Gibson, E. J., Shurcliff, A., & Yonas, A. A utilization of spelling patterns used by deaf and hearing subjects. In H. Levin & J. P. Williams (Eds.), *Basic studies on reading*. New York: Basic Books, 1970.

Henderson, L., & Chard, J. The reader's implicit knowledge of orthographic structure. In U. Frith (Ed.), *Cognitive processes in spelling*. New York: Academic Press, 1980.

Kucera, H., & Francis, W. N. *Computational analysis of present-day American English*. Providence: Brown University Press, 1967.

Manelis, L. The effect of meaningfulness in tachistoscopic word perception. *Perception & Psychophysics*, 1974, **16**, 182–192.

Massaro, D. W. Primary and secondary recognition in reading. In D. W. Massaro (Ed.), *Understanding language: An information processing analysis of speech perception, reading, and psycholinguistics*. New York: Academic Press, 1975.

Massaro, D. W. *Reading and listening* (Tech. Rep. No. 423). Madison: Wisconsin Research and Development Center for Individualized Schooling, December 1977.

Massaro, D. W. A stage model of reading and listening. *Visible Language*, 1978, **12**, 3–26.

Massaro, D. W. Letter information and orthographic context in word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, **5**, 595–609. (a)

Massaro, D. W. Reading and listening (Tutorial paper). In P. A. Kolers, M. Wrolstad, & H. Bouma (Eds.), *Processing of visible language* (Vol. 1). New York: Plenum, 1979. (b)

Massaro, D. W. How does orthographic structure facilitate reading? In J. F. Kavanagh & R. L. Venezky (Eds.), *Orthography, reading, and dyslexia*. Baltimore: University Park Press, 1980.

Massaro, D. W., Jastrzembski, J. E., & Lucas, P. A. *Frequency, orthographic regularity, and lexical status in letter and word perception* (Tech. Rep. No. 550). Madison: Wisconsin Research and Development Center for Individualized Schooling, 1980. (a)

Massaro, D. W., & Schmuller, J. Visual features, perceptual storage, and processing time in reading. In D. W. Massaro (Ed.), *Understanding language: An information processing analysis of speech perception, reading, and psycholinguistics*. New York: Academic Press, 1975.

Massaro, D. W., Taylor, G. A., Venezky, R. L., Jastrzembski, J. E., & Lucas, P. A. *Letter and word perception: Orthographic structure and visual processing in reading*. Amsterdam: North-Holland, 1980. (b)

Massaro, D. W., Venezky, R. L., & Taylor, G. A. Orthographic regularity, positional frequency, and visual processing of letter strings. *Journal of Experimental Psychology: General*, 1979, **108**, 107–124.

Mayzner, M. S., & Tresselt, M. E. Tables of single-letter and digram frequency counts for various word-length and letter-position combinations. *Psychonomic Monograph Supplements*, 1965, **1**, 13–32.

Mayzner, M. S., Tresselt, M. E., & Wolin, B. R. Tables of trigram frequency counts for various word-length and letter-position combinations. *Psychonomic Monograph Supplements,* 1965, **1,** 33–78.

McClelland, J. L., & Johnston, J. C. The role of familiar units in perception of words and nonwords. *Perception & Psychophysics,* 1977, **22,** 249–261.

Miller, G. A., Bruner, J. S., & Postman, L. Familiarity of letter sequences and tachistoscopic identification. *Journal of General Psychology,* 1954, **50,** 129–139.

Reicher, G. M. Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology,* 1969, **81,** 275–280.

Rosinski, R. R., & Wheeler, K. E. Children's use of orthographic structure in word discrimination. *Psychonomic Science,* 1972, **26,** 97–98.

Shannon, C. E. A mathematical theory of communication. *Bell System Technical Journal,* 1948, **27,** 379–423; 623–656.

Shannon, C. E. Prediction and entropy of printed English. *Bell System Technical Journal,* 1951, **30,** 50–64.

Shepard, R. N., & Podgorny, P. Cognitive processes that resemble perceptual processes. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes: Human information processing* (Vol. 5). Hillsdale, N.J.: Lawrence Erlbaum, 1978.

Solomon, R. L., & Postman, L. Frequency of usage as a determinant of recognition thresholds for words. *Journal of Experimental Psychology,* 1952, **43,** 195–201.

Solso, R. L., & King, J. F. Frequency and versatility of letters in the English language. *Behavior Research Methods & Instrumentation,* 1976, **8,** 283–286.

Spoehr, K. T. Phonological encoding in visual word recognition. *Journal of Verbal Learning and Verbal Behavior,* 1978, **17,** 127–141.

Taylor, G. A. *The contribution of orthographic structure to the perception of letter strings.* Unpublished doctoral dissertation, University of Kansas, 1977.

Taylor, G. A., Klitzke, D., & Massaro, D. W. Considerations in software character generation. *Behavior Research Methods & Instrumentation,* 1978, **10,** 787–788. (a)

Taylor, G. A., Klitzke, D., & Masssaro, D. W. A visual display system for reading and visual perception research. *Behavior Research Methods & Instrumentation,* 1978, **10,** 148–153. (b)

Taylor, M. M., & Creelman, C. D. PEST: Efficient estimates of probability functions. *Journal of the Acoustical Society of America,* 1967, **41,** 782–787.

Travers, J. R., & Olivier, D. G. Pronounceability and statistical "Englishness" as determinants of letter identification. *American Journal of Psychology,* 1978, **91,** 523–538.

Venezky, R. L., & Massaro, D. W. The role of orthographic regularity on word recognition. In L. Resnick & P. Weaver (Eds.), *Theory practice of early reading.* Hillsdale, N.J.: Lawrence Erlbaum, 1979.

Wheeler, D. D. Processes in word recognition. *Cognitive Psychology,* 1970, **1,** 59–85.