

Frequency of word usage and perceived word difficulty: Ratings of Kučera and Francis words

ALAN P. RUDELL

State University of New York, Health Science Center at Brooklyn, Brooklyn, New York

Twenty-four judges estimated the conceptual difficulty level of 870 five-letter words, using a 5-point scale. The words were selected from three word-frequency categories (1, 5-10, and 50-3,562/million) based on the word counts provided by Kučera and Francis (1967). The ratings were reliable. Tables in this paper list the means and standard deviations of the ratings for each word. Reaction time (RT) for valid word identification was tested in 20 subjects, using four sets of 50 words designed to test the effects of word frequency and word difficulty. RT was longer for more difficult words when word frequency was held constant. A word-frequency effect on RT was present when difficulty was held constant. The relationship of the results to subjective estimates of word familiarity is discussed.

The word-frequency effect is one of the best established findings in word recognition research. Typically, a subject's task is to distinguish valid words from nonwords as rapidly as possible, indicating a decision by a voluntary motor response. Reaction time (RT) for words that occur frequently in printed English tends to be shorter than for words that occur less often. Some word-frequency effect studies (Baker & Goodglass, 1979; Graf & Williams, 1987; Neisser & Beller, 1965) used Thorndike and Lorge's (1944) study to obtain word frequency data. Kučera and Francis's (1967) study was the source for others (Dobbs, Friedman, & Lloyd, 1985; Gerratt & Jones, 1987; Matlin & Derby, 1978; McCann, Besner, & Davelaar, 1988; Monsell, Doyle, & Haggard, 1989; Polich & Donchin, 1988; Sabol & DeRosa, 1976; Scarborough, Cortese, & Scarborough, 1977; Stuss, Picton, & Cerri, 1988).

The Kučera and Francis words were derived from an approximately one million word corpus of literature. Frequency of occurrence ranged from 1/million for many words to 69,971/million for the most commonly used word (THE). Examination of the counts given in their list shows that some of the words occurred as frequently as might be expected from a subjective standpoint. Other counts seemed to be jarringly inaccurate. For example, the word PIZZA occurred only three times, but the word HYMEN occurred 13 times.

Several explanations can be offered for the apparent discrepancy with everyday experience. At the time that the literature in the corpus was written, pizza may have been a less familiar item than it is today. Evolution of language might account for some discrepancies, but experimental results suggest other factors to be more important. Graf

and Williams (1987) asked subjects to complete three-letter stems to form words. The frequency of the first word completion supplied by the subjects was correlated significantly with language frequency as defined by both Kučera and Francis ($r = .30$) and Thorndike and Lorge ($r = .48$). The authors concluded that the Thorndike and Lorge norms provided a better estimate of word familiarity. This was surprising (Graf & Williams, 1987) because the Thorndike & Lorge norms were collected 23 years before the Kučera and Francis norms.

The Kučera and Francis literature corpus consisted of 500 samples of usually continuous passages, each about 2,000 words long. The limited number of topics treated and serial dependencies in the continuous passages probably resulted in certain words' occurring more often than in common usage. The three occurrences of PIZZA had three different sources, but the 13 occurrences of HYMEN all came from the same source. Perhaps the latter was from a gynecology journal. Even more striking, the word ANODE occurred 72 times, but in only 2 of the 500 samples. Since the word did not appear in any of the other sources, there was a highly disproportionate use of it in the two samples in which it did appear.

Serial dependency cannot explain all the apparently discrepant results. The word PHASE occurred 72 times and was found in 36 samples. SPITE occurred 56 times in 49 samples. The large number of sources for these words indicates that serial dependency was not an important factor. From a subjective standpoint, PHASE and SPITE do not seem much more common than words like JEWEL, GROAN, and SKATE. Yet the latter words occurred only once in the corpus. Kibby (1977) gives other anomalous examples. Many words used only once seem to be commonly known, even to a child. The words PUDDLE and DONKEY occurred only once per million, but they are considered at the 6-year-old level by intelligence tests (Wechsler Intelligence Scale for Children-Revised [WISC-R], and Stanford-Binet Intelligence Test).

This work was supported in part by Public Health Service Grant 1 R01 NS29340-01A1 awarded to the author. Correspondence should be addressed to A. P. Rudell, Department of Physiology, Box 31, SUNY, Health Science Center at Brooklyn, 450 Clarkson Ave., Brooklyn, NY 11203.

Since the literature in the corpus was written for publication, it seems reasonable to suppose that much of it was edited and revised to improve its quality. Slang expressions and words considered to be appropriate only when speaking to small children were probably avoided to maintain a professional tone. This could result in the apparent underrepresentation of some words commonly used in less formal language settings.

The considerations discussed above suggest that the word counts given by Kučera and Francis may have weaknesses as estimates of frequency of word usage by the general populace. If so, this would have important consequences for experiments that either study the word-frequency effect itself or use word-frequency estimates for balancing when studying the effects of other variables.

Even if the word counts did estimate frequency of word usage accurately, they might fail to accurately predict RT in certain cases. The word WHICH had a higher count than any other five-letter word (3,562 occurrences in 474 samples). It may have been used frequently because of its utility in sentence construction. It is doubtful that WHICH is conceptually that much easier than other words used less frequently, such as CHAIR (66/million) or APPLE (9/million).

Word difficulty is one of several variables related to the Kučera and Francis word counts. It can be defined in a number of ways, emphasizing any of a number of factors, such as degree of abstractness, specificity, or conceptual difficulty (Kibby, 1977). When difficulty was defined by intelligence test, it correlated with word frequency. Kibby reported a correlation of $-.58$ between Kučera and Francis frequency estimates and word difficulty as estimated from the WISC-R.

The effect of word frequency on RT performance is robust. The response time for correct identification is on the order of 50–100 msec longer for less frequently used words, depending on the magnitude of frequency difference. Differences of this order are easily detected and have consistently been found. Studies of other word features, such as bigram frequency composition and concreteness, yield less consistent results. A probable reason for the inconsistency was identified by Gernsbacher (1984). The experimental designs required balancing of word frequency, but the word counts used for this purpose were inadequate. They were particularly unreliable for low counts. The paradoxical results could be explained if a subjective variable, experiential familiarity, was used for predicting RTs. Schwanenflugel, Harnisfeger, and Stowe (1988) confirmed Gernsbacher's findings in the course of studying context availability.

Precedence in the learning of words might be an important factor for RT performance. Age-of-acquisition effects on RT have been obtained, but the relationship tends to vanish when word frequency, length, and familiarity have all been controlled (Gilhooly & Logie, 1982).

In the present study, adult judges were asked to estimate word difficulty. One objective was to estimate the degree of correspondence between perceived word difficulty and Kučera and Francis word-frequency counts. In

the definition of difficulty, emphasis was placed on the level of intellectual development required for use of a word. It was hoped that this definition would generate a statistic that could be used in conjunction with the word-frequency variable to provide substantially more accurate prediction of RT performance than could be obtained from word counts alone.

Word counts lack an index of variance. Individuals may disagree considerably about the difficulty of certain words. Variation in word difficulty across subjects is a source of error in many experiments. It decreases statistical power and generates less accurate estimates than otherwise might be obtained. The variance measure could be useful for excluding words with highly divergent ratings.

Some words (e.g., THUMB, QUIET), though commonly used and conceptually not difficult, even for a child, are harder to spell than are conceptually more difficult words. It did not seem reasonable to expect the subjects to be uninfluenced by such factors, even if instructed to ignore them. Neither were they asked to distinguish the concept of difficulty from related concepts, such as age of acquisition, familiarity, or perceived word frequency.

The main objective of the rating procedure was to obtain greater power for prediction of the recognition times of words used in electrophysiological experiments of visual recognition (Rudell, 1991, 1992). The brain wave response that occurs when a subject views a recognizable image is recorded in the presence of background noise. Response averaging is necessary to obtain an improved signal-to-noise ratio that is adequate for accurate estimation of the latency of the response. The response is biphasic. If the response latencies that comprise the average are highly variable, a latency estimate derived from that average may be inaccurate. Choosing word stimuli for an average that are recognizable with approximately equal ease should reduce latency variability and produce more accurate estimates of the latency of the electrophysiological response. The primary goal of this study, therefore, was to obtain, for a sufficiently large number of words, accurate estimates of speed of correct word identification. Accurate prediction of the latent period of a behavioral response was considered sufficient for this purpose, regardless of the specific mechanisms involved. Therefore, the success of the method could be judged solely by the power obtained for predicting RTs, whether this identified a single underlying variable or not.

METHOD

Difficulty Ratings

A set of 870 five-letter words was selected from those listed by Kučera and Francis. There were 290 words in each of three word-frequency categories. The high-usage category included words that occurred 50 or more times/million. Words occurring 5–10 times/million made up the medium-usage category. For the low-usage category, words were chosen that occurred once/million. On the basis of the generalization that words differing in frequency by a factor of 10 differ by about 50 msec in response time (Scarborough et al., 1977), the differences in frequency for the three word cate-

gories were expected to be large enough to detect significant differences in RT. Most of the five-letter words in a category were accepted, but proper names and plurals ending in "s" were avoided.

The 24 judges (14 males, 10 females) who rated the difficulty of the words were locally recruited volunteers (mean age = 42, range = 24–67 years). They had varied occupations: they were cafeteria workers, secretaries, technicians, nurses, graduate students, and college faculty members. The majority of the judges had bachelor's or higher university degrees. The large variation in age, profession, and education was deliberate. It was reasoned that differences in experience related to a person's education and vocation would result in words rated easy by some people and difficult by others. The standard deviation of the ratings should reflect such differences. Words with highly variable scores could then be excluded in subsequent experiments.

The judges received a randomly ordered list of the 870 words. They were asked to rate the difficulty of each word on a 5-point scale based on the level of development typically needed for use and comprehension by (1) a six-year old child, (2) a grade school student, (3) a high school student, (4) a well-read adult, or (5) a professor in his/her particular field of expertise. This criterion for difficulty was printed at the top of each page of the list.

The rating task was explained individually to each judge. Full use of the scoring range was encouraged. Questions about the task, such as the source of the words or the purpose of the rating procedure, were candidly answered. The judges were expected to perform the ratings during leisure moments. They were told to take as much time as they wanted for the task, but a response was required for every word. This procedure was designed to allow the judges to work at their own rate, the desirability of which was recognized by Gernsbacher (1984). It may have made hurried decisions even less likely than in the Gernsbacher procedure, because the task did not have to be completed in a single experimental session. All 24 of the scored word lists were returned within 2 months, with no need for prompting in most cases. Every word received a difficulty rating. Interviews with the judges after completion of the ratings showed that they had spent considerable time on the rating task, usually much more than 2 h. Many of them recalled how they had agonized over particular words. In general, their remarks reflected a genuine effort and suggested that they had tried hard to make accurate decisions.

For each judge, the ratings assigned to the 870 words were converted to standard scores. Consequently, the average score for each judge was 0 and the variance was 1.

Reaction Time Experiments

A computer algorithm generated four sets of 50 words for RT testing. They were constructed to permit testing of the effect of word frequency on RT when difficulty level was held constant and testing of the difficulty effect when word frequency was held constant. The program excluded words for which the standard deviation of the difficulty score was greater than 0.8. This rule disqualified the 39 words for which there was the least agreement about difficulty level. Words occurring more than 400 times/million were also excluded. This rule eliminated the 33 words with the highest word frequency and narrowed the range of frequencies permitted in a word set. The first set of words (E100) was expected to yield the shortest RTs. It was composed of words in the high-frequency category. The algorithm chose the 50 words that were rated least difficult and whose mean word frequency was closest to 100/million. The log of the word counts was used for these computations to correct for positive skew. The observed mean difficulty score for the 50 words was -1.16 ($SD = .06$). The observed mean word frequency was 99.9/million.

The second set of 50 words (A100) was also chosen from the high-frequency category. They were selected for mean frequency of 100/million and word difficulty of 0. Thus, these words were

of average difficulty, relative to the 870-word set. The observed mean difficulty score was 0.00 ($SD = .22$). The observed mean frequency was 100.5/million. RTs were expected to be longer for this set because the words were judged to be more difficult. Longer RTs for Set A100 could not be explained by the word-frequency effect, because the difference in word frequency was negligible, and the small difference observed favored shorter RTs for A100 than for E100.

A third set of words (A1) was chosen from the frequency category 1/million. A computer algorithm selected the words with the difficulty scores that most closely matched those of the words in Set A100. As a result, the observed mean difficulty was 0.00 ($SD = .22$), and the distribution of word difficulty for the two sets was very similar. If there was a word-frequency effect apart from word difficulty, RTs should be faster for Set A100 than for Set A1, since on the average the words of the former set occurred 100 times more frequently than did those of the latter.

The fourth set of words (D1) was chosen from the 1/million category. A target value for difficulty was selected that made the difference in difficulty ratings for the low-frequency words (0 vs. 1.16) the same as that for the high-frequency words (-1.16 vs. 0). The observed mean difficulty score was 1.16 ($SD = .20$).

People willing to serve without pay were locally recruited for the RT experiment. They were younger than the subjects who participated in word-difficulty rating. Older subjects were excluded because their RTs were likely to be longer and there was greater likelihood of problems with vision. Ten males (23–43 years, $M = 31.8$) and 10 females (22–38 years, $M = 28.5$) were tested. Every subject had some post-secondary education. All were right handed, and none reported uncorrected problems of vision or history of neurological disorder.

The subjects viewed a stream of random letters in which a valid target word was occasionally displayed (Rudell, 1992). The display device was a Zenith model ZVM-123, green-monochrome monitor. There were 3,200 distractors. They were five-letter strings generated by random selection of letters from the 200 experimental words. This produced equivalent letter frequency for target words and distractors. A judge screened the distractors, eliminating valid words and unusual sequences, such as all vowels or repetition of a letter three or more times. The target words were randomly mixed with the distractors. A subject signaled identification of a valid word by lifting the right index finger. This unblocked a beam of light that was detected by a phototransistor device. The subjects were asked to respond as quickly as possible, consistent with accurate performance.

Several practice viewing periods were permitted before the experimental words were presented. For this purpose, four sets of 50 words were chosen from the middle-frequency category. They averaged 7 occurrences/million. The mean difficulty ratings were -0.96 , -0.31 , 0.32 , and 0.97 . These words were presented only during practice trials to familiarize the subjects with the task and to ensure that they were able to detect a sufficiently large fraction of them, at least 50%. Correct word identification was usually well above 50% during practice, but one potential subject (who had previously apprised the experimenter of a problem referred to as "visual slowness") was unable to perform the task and was not included in the study.

The experimental words were tested in the same manner as the practice words. Five words from each experimental set were presented during a viewing period. Rest was permitted when the period ended. After 10 viewing periods, each of the 200 experimental words had been presented exactly once. The order of presentation was random. A different word order was used for each subject.

Statistical Analysis

The procedure described by Gernsbacher (1984) was used to correct the RTs. Responses in a word category that were more than

2.5 *SD* from the mean were replaced, following the winsorization method suggested by Winer (1971, pp. 51-54). A reciprocal transform converted the RTs to speed scores. Response speed was a variable that met the assumptions required for statistical analysis more closely than RT did.

The method of planned comparisons on repeated measures (Hays, 1963) was used to test the effect of word difficulty on RT. One comparison tested whether difficulty rating was an important factor for words occurring once/million. Another tested the difficulty effect for words that averaged 100 occurrences/million. A third comparison tested whether a word-frequency effect was present when word difficulty was held constant at the average value.

RESULTS

Interjudge Agreement

The mean difficulty scores for each word were computed separately for the 12 odd-numbered and 12 even-numbered subjects. A Pearson product-moment correlation coefficient (r), was calculated for odd and even groups, including the scores for all 870 words. The value of r was .96. This statistic showed that a substantial part of the variance in the mean judgments for one group was accounted for by the judgments made by the other group of subjects. Thus, the group judgments were in good agreement, indicating high reliability. Comparing the 12 oldest and the 12 youngest judges, r was .97. For males versus females, r was .96.

The reliability of a single judge was estimated by calculating r for his or her ratings and the mean ratings of the other 23 judges. The average value of r was .83 (range .62 to .90). Only three of the 24 judges had r values less than .80. Thus, the ratings of most judges conformed well to the group ratings.

Correspondence of Group Ratings to Word-Frequency Category

The group scores for the 870 words were correlated with frequency of usage, as defined by the three word-frequency categories. The value for r was .65; so frequency of usage accounted for a significant portion of the variance in mean estimates of word difficulty.

The 870 words were ranked by their mean difficulty scores. Each was assigned to one of three categories: the lowest, the middle, or the highest ranking. The cut-off point between the low to middle thirds was $-.533$ standard deviations. For the middle to high thirds, it was $.401$ standard deviations. For the 290 most commonly used words, 196, 89, and 5 words ranked in the lowest, middle, and highest third, respectively. For the medium word frequency, the corresponding numbers were 80, 121, and 89; for the least commonly used words, they were 14, 80, and 196. Thus, for the extreme word-frequency categories, about two thirds of the words received the expected ranking. Nearly one third of them were misplaced in the adjacent category. Far fewer words were misplaced in the more remote category. The middle third was unique in having two adjacent categories. This accounted for the smaller number of hits for the middle category than for

the extreme categories. For the middle frequency category, nearly one third of the responses were misplaced in the lower frequency categories and one third in the higher frequency categories; so less than half received the expected middle ranking.

Lists of the Words

The 870 words are listed in order of difficulty in Tables 1-3. Table 1 contains the words that occurred 50 or more times/million, Table 2 contains those occurring 5-10 times/million, and Table 3 contains those occurring once/million. The successive columns contain the word, the Kučera and Francis word count, mean difficulty score, and standard deviation of the difficulty score for the 24 judges. The means and standard deviations were multiplied by 100 to avoid decimal points.

In Table 1, the first 16 words all have the same mean score. This occurred because every judge assigned these words a score of 1. However, the standard deviation (.21) was not 0, because of the standardization procedure. This produced somewhat different scores for each individual's rating of 1, depending on the extent to which the full range of the 5-point scale was used. The mean difficulty score for Table 1 was $-.67$. The standard deviations averaged .42. All of the standard deviations were less than 1. For Table 2, the mean score was .01. The average standard deviation was .51. Three words (QUACK, CANON, and REPAY) stand out for having standard deviations greater than 1. For Table 3, the mean score was .67. The average standard deviation was .61. Eight words had standard deviations greater than 1. In general, there was somewhat greater variance in the scoring for words occurring less frequently.

Reaction Time Results

The rates of correct responding for word sets E100, A100, A1, and D1 were 87.1%, 81.2%, 69.9%, and 53.4%, respectively. False alarms occurred for 3.4% of the distractors. The word-difficulty effect was significant for the words that averaged 100 occurrences/million [$F(1,57) = 33.86, p < .01$]. RTs were longer for the more difficult words (585 msec) than for the easier words (556 msec). For words occurring once/million, word difficulty was also a significant factor [$F(1,57) = 59.36, p < .01$]. The response latency was 618 msec for the easier words and 668 msec for the more difficult ones. With word difficulty held constant at zero, a significant word frequency effect was present [$F(1,57) = 34.42, p < .01$]. The words had equal difficulty ratings, but the mean response latency for the 100/million words was 33 msec less than that for the 1/million words. These relationships are displayed in Figure 1.

Comparison With Familiarity Ratings

Gilhooly and Logie (1980) listed 1,944 words rated for familiarity and other variables. Of these, 145 were the same as those rated here for word difficulty. Word difficulty and word familiarity were negatively related ($r =$

Table 1
50 or More Occurrences/Million

WORD	KFWC	M	SD	WORD	KFWC	M	SD	WORD	KFWC	M	SD	WORD	KFWC	M	SD	WORD	KFWC	M	SD
BLACK	203	-123	21	BREAK	88	-107	35	LOVED	56	-87	49	AGREE	51	-64	39	GROSS	66	-31	52
CHAIR	66	-123	21	GUESS	56	-106	39	VISIT	109	-87	43	CHECK	88	-63	41	DAILY	122	-31	36
COLOR	141	-123	21	DRINK	82	-106	37	LIKED	58	-87	61	STUDY	246	-62	51	YOUTH	82	-30	36
GRASS	53	-123	21	RIVER	165	-106	34	BOARD	239	-86	45	APART	57	-62	52	OFFER	80	-28	38
HORSE	117	-123	21	DRIVE	105	-106	39	WORLD	787	-86	47	SOUTH	240	-61	58	SCENE	106	-27	36
HOUSE	591	-123	21	BROWN	176	-106	34	BLOCK	66	-85	51	BEGAN	312	-61	74	COURT	230	-26	26
LARGE	361	-123	21	START	154	-106	35	WHOLE	309	-85	44	BEGUN	51	-61	42	SENSE	311	-26	46
MOUTH	103	-123	21	FRONT	221	-105	34	OTHER	1702	-85	39	DROVE	62	-61	48	INNER	55	-26	58
PAPER	157	-123	21	CARRY	88	-105	32	TRUTH	126	-84	42	HEARD	247	-61	43	OUGHT	68	-24	52
PHONE	54	-123	21	WOMAN	224	-105	35	SPACE	184	-84	41	ABOUT	1815	-61	41	DOUBT	114	-22	54
QUIET	76	-123	21	AFTER	1070	-104	35	METAL	61	-84	52	WROTE	181	-60	43	WORTH	94	-22	47
SLEEP	65	-123	21	BEGIN	84	-104	33	EARLY	366	-83	41	THEIR	2670	-60	49	MEANT	100	-22	67
TABLE	198	-123	21	STONE	58	-103	39	FRESH	82	-83	41	FIFTY	68	-59	55	BRIEF	73	-22	47
TRUCK	57	-123	21	TOUCH	87	-103	43	READY	143	-82	43	ALLOW	72	-59	43	MINOR	58	-22	48
WATCH	81	-123	21	ROUND	81	-103	46	MOTOR	56	-82	47	WORRY	55	-58	44	AVOID	58	-21	49
WHITE	365	-123	21	SHORT	212	-102	41	THOSE	850	-82	50	ENEMY	88	-57	41	SHORE	61	-21	53
HAPPY	98	-120	28	PARTY	216	-102	37	CROWD	53	-81	43	BIRTH	66	-57	46	JUDGE	77	-21	40
WAGON	55	-120	26	CHAIN	50	-102	42	CLASS	207	-79	50	STATE	808	-57	44	ARMED	60	-19	40
WHEEL	56	-120	26	ABOVE	296	-102	37	STILL	782	-78	59	PROUD	50	-57	40	EVENT	81	-18	50
QUICK	68	-119	27	THESE	1573	-102	40	UNTIL	461	-78	48	TRUST	52	-57	40	STYLE	98	-18	30
FOUND	536	-118	26	COVER	88	-101	51	SPOKE	87	-77	40	ORDER	376	-56	36	LEGAL	72	-18	45
SMILE	58	-118	26	MONEY	265	-101	34	SHAPE	85	-77	49	STAGE	174	-53	47	APPLY	56	-15	41
TEETH	103	-118	26	PLANT	125	-101	38	TWICE	74	-76	42	FAVOR	78	-53	39	BASIC	171	-14	49
THREE	610	-118	26	WRITE	106	-101	35	OFTEN	368	-76	62	MODEL	77	-53	46	LEVEL	213	-14	32
WATER	442	-118	26	CHEST	53	-100	35	BUILT	103	-74	54	KNOWN	245	-52	52	MAJOR	247	-13	48
STAND	148	-118	28	MOVED	181	-100	38	ENTER	78	-74	48	SHOOK	57	-51	55	CLAIM	98	-10	37
EMPTY	64	-118	23	THIRD	190	-100	37	SPEND	53	-74	48	MIGHT	672	-51	60	DEPTH	53	-9	45
FLOOR	158	-118	23	ALONE	195	-98	42	ALONG	355	-74	46	WHOSE	252	-51	53	EXIST	59	-5	49
KNIFE	76	-118	23	STORE	74	-98	45	RIFLE	63	-73	60	SIGHT	86	-51	40	SPITE	56	-3	64
MUSIC	216	-118	23	BROKE	72	-97	42	COULD	1599	-73	50	SERVE	107	-50	40	RANGE	160	-2	41
NIGHT	411	-116	37	DREAM	64	-97	38	STOOD	212	-73	42	FRAME	74	-49	53	FAITH	111	-2	40
STORY	153	-116	37	GREAT	665	-97	40	EVERY	491	-72	49	LIVED	115	-49	54	IDEAL	61	-2	34
TRAIN	82	-116	37	MAYBE	134	-97	38	EARTH	150	-72	46	WHILE	680	-49	38	UNION	182	-1	46
SWEET	70	-115	29	SHARE	98	-96	41	CROSS	55	-72	52	SHOWN	166	-48	41	HONOR	66	-1	39
SMALL	542	-115	31	THINK	433	-95	47	EXTRA	50	-71	40	TRULY	57	-48	47	STAFF	113	2	43
RADIO	120	-114	28	HEAVY	110	-94	43	TAKEN	281	-71	44	BROAD	84	-47	48	ASIDE	67	3	50
UNCLE	57	-114	29	WHERE	938	-94	49	SCORE	66	-70	49	TOTAL	211	-47	48	VALUE	200	3	44
RIGHT	613	-114	26	FIGHT	98	-94	42	HOTEL	126	-70	41	GIVEN	377	-46	47	IMAGE	119	7	41
WRONG	129	-114	26	CLEAR	219	-94	43	THICK	67	-70	51	TITLE	77	-44	40	LABOR	149	7	49
EIGHT	104	-114	30	BELOW	145	-93	39	TRIED	170	-69	45	SPEED	83	-42	45	STOCK	147	8	41
SEVEN	113	-114	30	CLOSE	234	-93	38	CHIEF	119	-69	57	BEING	712	-42	55	AWARE	84	9	48
FIRST	1360	-113	29	NEVER	698	-93	40	RAISE	52	-69	46	FACED	54	-40	34	RURAL	54	9	55
GREEN	116	-113	39	LOWER	123	-93	39	ADDED	172	-69	48	TRADE	143	-40	51	ISSUE	152	11	49
TASTE	59	-113	39	PRICE	108	-92	39	WOULD	2714	-69	49	FORCE	230	-40	42	NOVEL	59	16	49
DRESS	67	-111	40	YOUNG	385	-92	37	FIELD	274	-68	49	DOZEN	52	-39	47	FORTH	71	17	60
UNDER	707	-111	35	THING	333	-92	39	DEATH	277	-68	44	NORTH	206	-39	48	MORAL	142	19	42
SPEAK	110	-111	36	VOICE	226	-92	41	SPENT	104	-68	45	SOLID	77	-39	37	BASIS	184	24	50
THERE	2724	-110	36	ALIVE	57	-91	49	AHEAD	109	-67	51	NEGRO	104	-38	36	FRANK	68	24	59
CLEAN	70	-110	29	HEART	173	-91	52	WHICH	3562	-67	54	CAUSE	130	-38	67	NOTED	90	26	47
DANCE	90	-110	31	PLANE	114	-91	79	SHALL	267	-66	57	FINAL	156	-36	45	UNITY	71	28	60
CHILD	213	-110	36	LATER	397	-90	47	QUITE	281	-66	52	DRAWN	70	-35	52	BASED	119	29	49
BRING	158	-109	32	AGAIN	578	-90	40	PEACE	198	-65	50	POWER	342	-34	45	INDEX	81	37	50
TODAY	284	-109	29	WOMEN	195	-90	40	SINCE	628	-65	48	FULLY	80	-32	55	THEME	55	38	48
SHARP	72	-109	30	REACH	106	-89	41	SCALE	60	-65	47	AMONG	370	-32	39	VITAL	56	40	40
BLOOD	121	-109	31	SOUND	204	-89	49	PRESS	127	-65	52	USUAL	96	-32	36	CIVIL	91	54	46
LIGHT	333	-108	42	BUILD	86	-88	44	LOOSE	53	-64	45	TRIAL	134	-32	35	PHASE	72	54	58
BEACH	61	-108	42	MONTH	130	-88	42	UPPER	72	-64	50	PROVE	53	-31	41	HENCE	58	62	43
LEAVE	205	-107	38	PIECE	129	-88	46	MARCH	120	-64	47	FLESH	52	-31	47	ANODE	77	113	65

Note—The words are listed in order of difficulty under WORD. KFWC is the Kučera and Francis (1967) word count. *M* is the mean difficulty score for 24 judges. *SD* is the standard deviation of the difficulty score. Means and standard deviations were multiplied by 100 to avoid decimal points.

Table 2
5-10 Occurrences/Million

WORD	KFWC	M	SD	WORD	KFWC	M	SD	WORD	KFWC	M	SD	WORD	KFWC	M	SD	WORD	KFWC	M	SD
APPLE	9	-123	21	NIECE	8	-68	48	SCOUT	8	-22	49	KNELT	8	13	64	ENACT	7	70	39
MOUSE	10	-123	21	TYING	5	-67	47	BLEND	9	-22	38	DONOR	5	14	41	LOFTY	5	71	43
SPOON	6	-123	21	SMOKY	5	-67	59	VOWEL	7	-22	50	ARENA	7	15	62	TORSO	7	71	50
WITCH	5	-123	21	ALBUM	6	-66	50	HITCH	5	-22	43	NOTCH	6	16	48	STARK	7	71	35
GLOVE	9	-119	25	CHORE	7	-66	55	GROOM	5	-22	47	GRIEF	10	16	46	ASSET	5	72	29
KITTY	7	-119	35	TEASE	6	-63	46	QUACK	9	-22	102	TIMID	5	17	59	ABBEY	7	72	64
TIGER	7	-118	26	BURNT	6	-63	48	WIDEN	5	-21	42	HUNCH	7	19	49	SIEGE	6	75	41
THUMB	10	-118	23	FATTY	7	-63	55	SHONE	5	-21	60	GAZED	7	20	43	FORUM	10	75	41
HELLO	10	-116	37	STOLE	10	-62	49	LILLY	10	-20	63	LEASE	10	20	51	FOLLY	10	75	44
BACON	10	-114	28	ROCKY	10	-62	46	LATCH	5	-19	53	WITTY	10	22	55	PLUMB	5	76	86
ANKLE	8	-109	32	PHOTO	5	-61	55	OTTER	5	-19	66	SLACK	9	25	41	SONAR	7	77	60
SHOUT	9	-108	43	BEAST	7	-61	59	SATIN	5	-15	32	TRUCE	5	27	40	RESIN	9	78	44
TOWEL	6	-108	39	THIEF	8	-61	43	SPECK	7	-15	48	FRAIL	8	29	47	ROTOR	6	78	54
ELBOW	10	-107	35	COMIC	9	-60	49	CABLE	7	-14	37	SEIZE	6	29	46	SPIRE	5	78	66
COUGH	7	-106	36	HOUND	7	-59	45	THEFT	10	-14	48	RALLY	10	29	50	BOGEY	5	78	74
SALAD	9	-106	33	OLIVE	5	-58	50	LINEN	6	-14	46	FILLY	9	30	67	FORTE	6	79	38
NOISY	6	-105	39	TENTH	7	-57	50	CANON	5	-13	104	TREAD	5	30	68	CREED	8	79	46
SNACK	6	-105	36	WRECK	8	-57	49	WISER	7	-10	50	DERBY	7	31	59	LADEN	6	79	43
SHAVE	6	-103	36	FRIED	6	-56	49	AISLE	6	-10	58	HASTY	5	32	48	RELIC	6	79	43
STEAK	10	-98	36	CHILI	6	-55	54	BATCH	5	-10	58	MANOR	5	33	61	BLOAT	8	80	31
EAGLE	5	-98	44	STAIN	6	-55	43	TOPIC	9	-9	42	TWEED	5	34	57	STINT	6	80	50
STAMP	8	-97	39	CANOE	7	-54	40	HASTE	9	-9	46	DREAD	9	35	47	PAPAL	7	81	51
STING	5	-97	38	PEARL	9	-53	44	PULSE	9	-9	60	LUNAR	10	36	59	PIOUS	10	81	72
PUSSY	5	-94	40	MINUS	8	-53	51	BITCH	6	-9	57	ALIBI	8	36	45	VOGUE	6	83	36
TRUNK	8	-94	52	ALLEY	8	-51	52	BLAZE	7	-7	56	FLASK	5	37	46	GAUNT	6	83	32
JUICY	6	-93	40	FUZZY	7	-50	54	REPAY	7	-7	120	DWELL	8	39	57	ALOOF	5	84	32
DIZZY	5	-92	41	STOOL	8	-47	66	ELECT	8	-7	39	GULLY	5	40	61	TITAN	7	85	58
CURLY	5	-91	43	SCOOP	5	-47	52	TOKEN	10	-7	51	ALGAE	7	42	67	INCUR	5	86	48
PUNCH	5	-91	47	FETCH	6	-45	63	UNITE	10	-6	40	GLAND	9	43	62	REBUT	6	86	44
MUDDY	10	-91	48	SEWER	10	-44	50	SPINE	6	-5	49	TENOR	6	43	49	INERT	5	87	40
APRON	7	-89	42	MEDAL	7	-44	60	PANSY	6	-5	41	MERGE	10	43	48	EVOKE	6	90	52
BERRY	9	-89	43	ROAST	10	-43	41	MOODY	5	-5	51	REPEL	8	46	53	GUILD	7	91	43
CIGAR	10	-88	38	WHEAT	9	-41	49	POKER	6	-3	35	VOWED	5	47	49	MOTIF	8	96	33
SUPER	8	-87	46	CHEER	8	-40	40	BASIN	7	-3	66	BOSOM	8	48	60	GUISE	6	97	41
FROZE	5	-86	45	SAFER	5	-40	47	ASIAN	10	-2	50	MURKY	5	49	64	PROXY	7	100	41
PINCH	6	-86	49	SWELL	7	-39	41	SPADE	10	0	57	FRAUD	8	50	40	BEECH	6	101	57
STEAL	5	-84	40	RINSE	6	-38	57	PLANK	7	1	60	FIERY	7	51	88	IONIC	8	101	49
SWORD	7	-83	52	GREET	7	-37	65	DINGY	5	3	58	ADAPT	5	51	39	BESET	7	104	61
WRIST	10	-83	52	CRANE	5	-35	38	SWORN	5	3	44	GAILY	5	54	70	MULCH	6	105	51
RAINY	5	-82	54	RUSTY	8	-35	49	POLAR	7	3	62	SURGE	9	56	52	GAMMA	5	105	50
PASTE	10	-82	61	THIGH	9	-35	67	METER	6	3	61	MANIA	5	57	46	FREER	5	107	91
SANDY	6	-81	48	SCENT	6	-34	55	CLING	6	3	54	PERIL	8	57	43	TONAL	9	114	50
AWOKE	9	-80	51	SHAKY	5	-33	61	CORAL	5	4	53	VILLA	6	58	56	ARGON	7	117	50
CHOKE	9	-79	51	HATCH	5	-31	61	BRUTE	6	4	43	PROBE	6	58	37	EPOCH	6	118	73
SCRUB	9	-78	62	CHOIR	8	-31	45	LEDGE	6	4	60	ROUGE	7	59	78	DIVAN	6	119	70
BADGE	5	-77	51	EASEL	5	-28	54	BATON	10	5	59	LAPSE	6	60	44	SKIFF	9	119	74
SWEAR	10	-76	43	MAPLE	7	-28	31	CHORD	7	5	46	FLAIR	8	60	53	IDIOM	7	123	63
FROST	6	-75	49	BARGE	7	-28	33	ALTAR	5	6	65	FAMED	5	61	61	CREDO	8	124	60
WEIRD	10	-74	60	DITCH	10	-27	38	BULGE	5	6	39	WRATH	9	61	47	STEAD	5	126	75
KNEEL	5	-73	59	SWAMP	5	-27	44	SLUMP	8	7	48	CAPER	6	62	60	SERGE	5	137	82
FLOUR	8	-72	46	DEMON	9	-24	59	BLUFF	8	8	42	NAIVE	7	62	37	RUPEE	6	144	90
FOGGY	5	-72	50	CREEP	10	-24	34	POSED	7	8	55	SCANT	5	63	37	SEPTA	6	155	62
SHINE	5	-72	41	PAVED	5	-23	50	SLATE	10	8	67	DELTA	7	63	53	SERVO	5	160	87
MERRY	8	-72	53	QUEER	6	-23	55	RUMOR	8	10	45	FORGE	10	63	49	MORES	7	162	82
BUGGY	6	-70	56	FLOCK	10	-23	47	AWAIT	9	11	54	BLEAK	10	63	48	HILUM	5	181	73
NASTY	5	-69	47	SQUAT	7	-23	53	CLASH	5	11	41	BROOD	9	66	50	QUIRT	8	195	91
SPEAR	7	-69	48	BOOTH	7	-23	49	AGONY	9	12	44	POISE	6	66	35	SITUS	5	228	72
STACK	9	-68	45	STONY	5	-22	77	PLEAD	5	13	52	QUILL	9	70	55	LEMMA	7	231	55

Note—The words are listed in order of difficulty under WORD. KFWC is the Kučera and Francis (1967) word count. *M* is the mean difficulty score for 24 judges. *SD* is the standard deviation of the difficulty score. Means and standard deviations were multiplied by 100 to avoid decimal points.

Table 3
1 Occurrence/Million

WORD	KFWC	M	SD	WORD	KFWC	M	SD	WORD	KFWC	M	SD	WORD	KFWC	M	SD	WORD	KFWC	M	SD
BUNNY	1	-123	21	GHOUL	1	-3	69	ANTIC	1	54	56	CLOUT	1	83	43	ACRID	1	127	67
SKATE	1	-116	37	POLKA	1	-2	46	ANNEX	1	54	48	LIVID	1	83	32	STAIN	1	129	52
ZEBRA	1	-110	36	SHOWY	1	-1	33	STEED	1	55	57	GIRTH	1	84	46	ELFIN	1	129	98
SPILL	1	-107	41	SQUAW	1	1	74	DITTY	1	58	72	RANDY	1	85	108	LEARY	1	129	98
CAMEL	1	-94	45	PRANK	1	2	44	BICEP	1	58	61	ROGUE	1	86	41	AXIOM	1	130	62
CRUST	1	-92	46	CREAK	1	3	51	GLOSS	1	58	44	EXALT	1	87	49	WREST	1	132	84
SNAIL	1	-84	51	PLAID	1	4	56	YEARN	1	58	54	FLORA	1	87	55	BORON	1	132	65
ICING	1	-81	49	UNFIT	1	6	47	MINCE	1	59	43	WETLY	1	87	118	LUMEN	1	136	69
BLIMP	1	-68	54	GAUZE	1	6	41	SINUS	1	60	54	LATHE	1	87	52	SKULK	1	136	73
BOXER	1	-66	48	UNCAP	1	8	57	POSEY	1	61	133	JOUST	1	87	59	FORAY	1	137	62
ROBOT	1	-65	64	MOLAR	1	9	65	TARRY	1	62	75	BRIAR	1	88	46	PLASH	1	138	82
JEWEL	1	-64	43	LADLE	1	11	67	VIGIL	1	62	50	GROPE	1	88	51	MITER	1	139	49
FLUTE	1	-58	58	CADDY	1	11	49	WIELD	1	62	50	TOXIN	1	89	47	DINGO	1	143	113
CHAMP	1	-58	53	SCALD	1	14	65	SIEVE	1	62	55	ARBOR	1	89	52	CAROB	1	143	80
FLAKE	1	-44	53	WORDY	1	15	48	EMCEE	1	63	76	SAUTE	1	90	62	ILEUM	1	144	64
GROAN	1	-43	49	BRIBE	1	15	59	WRACK	1	63	91	LEVEE	1	90	50	DATUM	1	145	68
SHACK	1	-43	45	PERCH	1	16	49	EVADE	1	65	54	BLEAT	1	90	83	DRUID	1	145	61
SPICY	1	-42	49	BRAWL	1	16	49	BRUNT	1	65	68	TRUMP	1	91	56	RAJAH	1	146	86
CORNY	1	-40	43	SAUCY	1	20	51	VALOR	1	66	45	COPLY	1	91	37	SHILL	1	146	88
CAGED	1	-39	59	DIGIT	1	23	62	CHIVE	1	66	42	CACHE	1	92	58	PARRY	1	150	61
TAFFY	1	-39	60	HYENA	1	24	76	TEPID	1	66	38	AFOOT	1	92	45	ANION	1	150	81
MUNCH	1	-37	75	CEDAR	1	24	52	TONIC	1	67	51	MOULD	1	93	90	PITHY	1	151	57
HIKED	1	-36	36	TIDAL	1	24	50	COCAO	1	67	149	KEBOB	1	93	95	AURAL	1	151	60
FAKER	1	-36	67	TANGY	1	25	58	BINGE	1	67	40	QUERY	1	93	41	SEPIA	1	154	78
SHADY	1	-35	42	MOUSY	1	26	73	CRYPT	1	67	61	SLOOP	1	94	64	PHYLA	1	155	77
WORMY	1	-33	61	KAZOO	1	27	95	HERON	1	67	77	FORGO	1	95	68	AMIDE	1	157	60
SIREN	1	-29	76	EQUIP	1	27	44	FELON	1	68	39	OMEGA	1	95	35	SCION	1	157	61
WHACK	1	-29	55	LAPEL	1	29	49	TUNIC	1	68	51	DEITY	1	97	51	COCAO	1	158	96
SPOUT	1	-26	46	WHIFF	1	30	57	FROTH	1	69	55	MAUVE	1	97	57	POESY	1	163	94
LEAFY	1	-26	75	RARER	1	31	61	BELLE	1	70	64	ATUNE	1	97	55	ILIAC	1	164	72
ROOMY	1	-24	42	CLEAT	1	33	62	BUXOM	1	70	44	SANER	1	98	68	ECOLE	1	165	81
PECAN	1	-23	55	BEEFY	1	34	56	NYMPH	1	71	40	SWATH	1	99	60	BOURN	1	167	90
SLEET	1	-23	59	SHAKI	1	34	55	PENAL	1	72	37	HOAGY	1	102	96	SIDLE	1	175	88
URINE	1	-21	64	SNEER	1	36	42	AFFIX	1	72	33	OCTET	1	103	54	ELEGY	1	176	61
TRAMP	1	-21	54	ANVIL	1	37	66	INFER	1	73	47	GLEAN	1	104	46	INAPT	1	179	111
DROOP	1	-19	47	SALON	1	39	56	TORAH	1	73	66	DETER	1	104	100	KIOSK	1	182	90
MINER	1	-18	40	ROOST	1	40	58	GRAFT	1	73	50	PORGY	1	106	95	SERIF	1	186	65
CROAK	1	-18	49	EXCEL	1	40	58	DWELT	1	74	54	SHUNT	1	106	55	FUSTY	1	189	81
SCUFF	1	-18	39	JAZZY	1	41	41	ATONE	1	74	37	DAUNT	1	107	47	OHMIC	1	190	86
NYLON	1	-18	35	FILMY	1	41	45	CHAFE	1	76	50	FLAIL	1	110	68	VELDT	1	191	74
GENIE	1	-17	68	GNOME	1	44	73	AVERT	1	76	35	MAXIM	1	112	61	SYNOD	1	193	70
GRAZE	1	-15	33	WREAK	1	44	71	IRATE	1	76	38	AMINO	1	114	57	COZEN	1	194	91
PUTTY	1	-15	61	FUMED	1	45	62	SUMAC	1	76	68	TRIAD	1	114	65	AEGIS	1	199	74
PRUNE	1	-14	62	SNARE	1	45	49	ETHER	1	76	62	NATTY	1	114	70	ARGOT	1	199	76
BEIGE	1	-13	62	KNEAD	1	45	51	QUIRK	1	76	39	WAXEN	1	115	64	TANIN	1	199	75
PATTY	1	-13	62	SABER	1	46	60	PACER	1	77	53	GIMPY	1	115	61	SCRIM	1	205	95
YODEL	1	-13	49	RUMMY	1	46	86	TRIBE	1	77	51	CILIA	1	116	63	PREXY	1	205	79
FLUFF	1	-12	53	ALIAS	1	47	51	SAVVY	1	78	38	GUILE	1	118	52	ECLAT	1	206	67
HUMID	1	-10	56	WAIVE	1	47	85	AWASH	1	79	55	SAMBA	1	119	63	MOTET	1	208	76
RODEO	1	-10	69	FLUKE	1	47	49	SPRIG	1	80	42	FAUNA	1	120	69	NONCE	1	209	70
MUGGY	1	-10	51	GAWKY	1	47	52	FRIAR	1	80	79	INANE	1	120	62	PLAYA	1	214	133
SCOUR	1	-9	47	GASSY	1	47	54	RETCH	1	81	76	HARPY	1	120	86	RHEUM	1	217	89
SILKY	1	-8	43	HOIST	1	48	59	MECCA	1	81	41	RENAL	1	120	49	MANSE	1	217	53
FLIRT	1	-7	54	AFIRE	1	49	56	SHANK	1	81	52	USURP	1	121	65	MASER	1	222	76
SNOUT	1	-6	52	WHOOPE	1	51	83	FRANC	1	81	72	AMITY	1	122	49	LISLE	1	223	77
WALTZ	1	-6	34	BASIL	1	53	58	BRASH	1	82	53	BUTTE	1	123	58	RIVEN	1	226	63
HINGE	1	-5	57	GOUGE	1	53	38	SWANK	1	82	43	XENON	1	125	70	FRISE	1	240	78
SHRUB	1	-4	58	SPOOF	1	54	47	CREPE	1	82	47	LUGER	1	125	68	TILTH	1	258	82

Note—The words are listed in order of difficulty under WORD. KFWC is the Kučera and Francis (1967) word count. *M* is the mean difficulty score for 24 judges. *SD* is the standard deviation of the difficulty score. Means and standard deviations were multiplied by 100 to avoid decimal points.

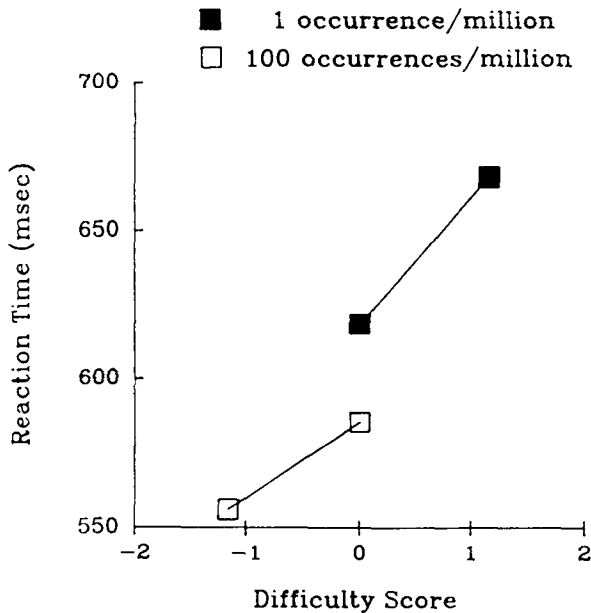


Figure 1. Reaction time and word difficulty. The mean RT for 20 subjects is plotted for four word sets. Sets E100 and A100 (open squares) contained words that averaged 100 occurrences/million. The words in Sets A1 and D1 (filled squares) occurred once/million.

-.77). The majority of these words were in the higher frequency category. Of these, 27 were tested in the RT experiment. Both ratings were significantly related to response speed. For word difficulty, r was $-.55$. For familiarity, r was $.45$. The difficulty and familiarity scores were significantly related for the 27 words ($r = -.70$). When familiarity was held constant, the partial correlation coefficient for response speed and difficulty was $-.36$. This value was statistically significant ($p < .05$). With word difficulty held constant, the partial correlation coefficient of response speed with familiarity was not significant ($r = .12$, $p > .05$). The same pattern of results was observed for the other variables provided by Gilhooly and Logie. For age of acquisition, imagery, concreteness, log Thorndike and Lorge word count, and log Kučera and Francis word count, the partial correlation for response speed with difficulty was significant when these other variables were held constant. None of these correlations were significant when word difficulty was held constant.

The number of words in the low-frequency category that were tested in the RT experiment and were common to the Gilhooly and Logie words was insufficient for statistical analysis. However, Gernsbacher (1983) analyzed 455 low-frequency words that were rated for familiarity.¹ Of these, 75 were common to those judged for difficulty. Word difficulty and familiarity were negatively related ($r = -.81$). Of the 75 common words, 19 were tested in the RT experiment. Both word difficulty ($r = -.56$) and familiarity ($r = .49$) were significantly related to re-

sponse speed. The correlation between difficulty and familiarity for the 19 words was $-.85$. With familiarity held constant, the partial correlation coefficient of response speed with word difficulty was $-.30$, not significant at the $.05$ level. With difficulty held constant, the partial coefficient of response speed with familiarity was insignificant ($r = .05$, $p > .05$).

Of the words scored for familiarity in the Gilhooly and Logie and the Gernsbacher studies, 33 were the same. The familiarity scores were positively correlated ($r = .69$).

DISCUSSION

The results show that the word-difficulty ratings were reliable and, like word frequency, had predictive power for speed of correct word identification. The question of whether word-frequency or word-difficulty rating has more impact on performance is meaningless unless some equivalence of units can be established. In this study, a difference of 1.16 in difficulty score with frequency held constant was approximately equal to a 100-to-1 difference in word frequency when difficulty was held constant. The observed differences in RT were of the same order as those reported for the word-frequency effect in other studies (Dobbs et al., 1985; Gerratt & Jones, 1987; Monsell et al., 1989; Scarborough et al., 1977).

The words tested in the RT experiment included those that had the lowest word-frequency value possible for Kučera and Francis words, 1/million. The greatest difficulty score for any word tested was 1.51. There were 35 words in the 1/million category that had greater difficulty scores, ranging from 1.54 to 2.58. A linear relationship of difficulty score to response speed cannot be assumed, but it seems reasonable to suppose that response times would have been even slower for these words, if tested, than for those that averaged 1.16. Thus, the range of difficulty score tested could be increased by including words with larger difficulty scores, whereas the Kučera and Francis word count had reached its limit at 1/million.

The results showed that word difficulty ratings were at least as effective as familiarity ratings for predicting response time. There was some evidence that difficulty was a better predictor than was familiarity. This conclusion should be tempered by the following consideration. The subjects who participated in the RT experiments were drawn from the same geographic area as those who rated the words. Actual word usage may differ significantly from one region to another. Had the subjects been drawn from Scotland or Texas instead of New York, the apparent superiority of the difficulty ratings might have disappeared if it depended primarily on regional differences in word usage.

Whether or not difficulty is a better predictor of RT than is familiarity, there is little doubt that consideration of either variable in conjunction with word frequency improves prediction of RT over that derived from consideration of word frequency alone. If it is assumed that difficulty and familiarity are significantly different factors, an

investigator may wish to balance some words by familiarity, others by difficulty, to show that neither variable could explain the experimental effects that were obtained.

Investigators faced with a need to balance both word frequency and familiarity may find the available pool of rated five-letter words too small for their purposes. This is particularly a problem if word length must be held constant. If it can be assumed that for balancing purposes the familiarity ratings that were done in Scotland are equivalent to those that were done in Texas and to the difficulty ratings that were done in New York, then the tables provided here greatly increase the available number of rated words. Only 145 words were included among the 348 five-letter words given in Gilhooly and Logie (1980), yielding 725 additional rated words. There were 795 words that were not included among the 445 rated for familiarity in Gernsbacher's (1983) study; so 75 of them overlapped. The number of overlapping words is probably sufficient to establish an equivalence for familiarity and difficulty ratings. A statistic could be generated that applied to a larger number of five-letter words.

To decrease the amount of computation required, the familiarity and difficulty scores could be left in the units given, balancing some word categories by one, others by another type of score. If the assumption of equivalent balancing effectiveness is correct, a statistical test should reject the hypothesis that the experimental effect depended on the score used for balancing.

The improvement in predictive power produced by taking word difficulty into account in addition to word frequency was not restricted to words that occur rarely. Words occurring 100 times/million on the average showed significant differences in response time that depended on word difficulty. Thus, analysis of both high- and low-frequency words could benefit from consideration of rated word difficulty.

REFERENCES

- BAKER, E., & GOODGLASS, H. (1979). Time for auditory processing of object names by aphasics. *Brain & Language*, **8**, 355-366.
- DOBBS, A. R., FRIEDMAN, A., & LLOYD, J. (1985). Frequency effects in lexical decisions: A test of the verification model. *Journal of Experimental Psychology: Human Perception & Performance*, **11**, 81-92.
- GERNSBACHER, M. A. (1983, April). *The experiential familiarity norms and their psychological reality*. Paper presented at the 29th Annual Meeting of the Southwestern Psychological Association, San Antonio, TX.
- GERNSBACHER, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, **113**, 256-281.
- GERRATT, B. R., & JONES, D. (1987). Aphasic performance on a lexical decision task: Multiple meanings and word frequency. *Brain & Language*, **30**, 106-115.
- GILHOOLY, K. J., & LOGIE, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation*, **12**, 395-427.
- GILHOOLY, K. J., & LOGIE, R. H. (1982). Word age-of-acquisition and lexical decision making. *Acta Psychologica*, **50**, 21-34.
- GRAF, P., & WILLIAMS, D. (1987). Completion norms for 40 three-letter word stems. *Behavioral Research Methods, Instruments, & Computers*, **19**, 422-445.
- HAYS, W. L. (1963). *Statistics for psychologists*. New York: Holt Rinehart & Winston.
- KIBBY, M. W. (1977). Note on relationship of word difficulty and word frequency. *Psychological Reports*, **41**, 12-14.
- KUČERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- MATLIN, M. W., & DERBY, P. L. (1978). Relationship between repression-sensitization and word frequency estimates for pleasant and unpleasant words. *Perceptual & Motor Skills*, **46**, 351-354.
- MCCANN, R. S., BESNER, D., & DAVELAAR, E. (1988). Word recognition and identification: Do word-frequency effects reflect lexical access? *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 693-706.
- MONSELL, S., DOYLE, M. C., & HAGGARD, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General*, **118**, 43-71.
- NEISSER, U., & BELLER, H. K. (1965). Searching through word lists. *British Journal of Psychology*, **56**, 349-358.
- POLICH, J., & DONCHIN, E. (1988). P300 and the word frequency effect. *Electroencephalography & Clinical Neurophysiology*, **70**, 33-45.
- RUDELL, A. P. (1991). The recognition potential contrasted with the P300. *International Journal of Neuroscience*, **60**, 85-111.
- RUDELL, A. P. (1992). Rapid stream stimulation and the recognition potential. *Electroencephalography & Clinical Neurophysiology*, **83**, 77-82.
- SABOL, M. A., & DE ROSA, D. V. (1976). Semantic encoding of isolated words. *Journal of Experimental Psychology: Human Learning & Memory*, **2**, 58-68.
- SCARBOROUGH, D. L., CORTESE, C., & SCARBOROUGH, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception & Performance*, **3**, 1-17.
- SCHWANENFLUGEL, P. J., HARNISHFEGER, K. K., & STOWE, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory & Language*, **27**, 499-520.
- STUSS, D. T., PICTON, T. W., & CERRI, A. M. (1988). Electrophysiological manifestations of typicality judgment. *Brain & Language*, **33**, 260-272.
- THORNDIKE, E. L., & LORGE, I. (1944). *The teacher's word book of 30,000 words*. New York: Bureau of Publications, Teachers College, Columbia University.
- WINER, B. J. (1971). *Statistical principles in experimental design*. New York: McGraw-Hill.

NOTE

1. I am grateful to Morton Ann Gernsbacher for supplying me with a list of the words and familiarity ratings.

(Manuscript received June 6, 1991;
revision accepted for publication June 7, 1993.)