Jaap Phil Nacke (ed) 23rd Yearbook of the National Reading Conference (pp 176-184). Clemson, SC: National Reading Conference.

HOW DOES VOCABULARY ACCOUNT FOR VARIANCE ON READING COMPREHENSION TESTS? A PRELIMINARY INSTRUCTIONAL ANALYSIS*

J. JAAP TUINMAN
MARY ELLA BRADY
Indiana University

The simple question that motivated us to embark on a program of research, in which the study reported here is a first exploration, is: Given the fact that vocabulary knowledge accounts for a sizeable amount of variance on reading comprehension tests (by anyone's accounting), does this mean that instruction in vocabulary will enhance performance on comprehension tests?

Of course this question, which is basically one of those posed by Davis (1971) when he considered next steps in research in reading comprehension, is not as simple as it may appear. First, it is part of a larger group of questions regarding the adequacy of using logical and/or empirical analyses of reading behavior in the formulation of component skills in which children need instruction. Secondly, expectations that the results of vocabulary instruction transfer to the tasks involved in measures of reading comprehension seem to be justified only if one of the following three conditions applies: (1) the vocabulary taught makes up the reading material to be comprehended, (2) the vocabulary taught is a sample from a rather limited population of words, which together constitute the body of language which makes up the reading material to be comprehended, and (3) the instructional program in vocabulary focuses on generalized vocabulary skills, such as semantic-developmental variables. Thus, emphasis is not on the learning of specific word meanings, but on the acquisition of refined and extended semantic structures. Two recent reviews of semantic development in young children (Palermo & Molfese, 1972; Perfetti, 1972) make clear that we may not know exactly what we mean by instructional goals such as "refined and extended semantic structures." Nevertheless, such an objective has intuitive appeal and may be a logical *a priori* for vocabulary instruction on a large scale.

The present study deals with the first of the three conditions stated above. The specific question asked in this study was: To what extent will performance on a reading comprehension test improve as a function of direct instruction in the meaning of all difficult words occurring in the paragraphs used in the test?

Why ask this particular question? We frequently say, "Vocabulary

*The authors wish to thank Dr. D. Houston and his faculty at the Hoover Elementary School for their marvelous cooperation.

accounts for a large part of the variance on comprehension measures." The interpretation of this statement normally is either statistical or psychological. In the former case, it becomes a statement of mathematical fact; in the latter case, the statement asserts that a trait can be hypothesized which explains shared variance on both the vocabulary and comprehension tests.

Traits, however, are hypothetical and generally not too amenable to instruction. Yet, the "account for" image is frequently intended as a meaningful assertion in instructional (as opposed to statistical and psychological) discourse. But what do we mean in that case? Do we mean that we must know the meanings of the words in a passage before that passage can be understood? Do we mean that vocabulary knowledge results in semantic sensitivity (in some definable sense) which in turn results in high comprehension scores? Do we mean that for children with large vocabularies the probability of encountering prose with too many unknown words is relatively small and that, therefore, they tend to score well on comprehension tests?

The questions asked here can be both extended and refined, but for the present purpose they serve to illustrate our thesis: It is presently not at all clear which educational interpretations of the assertion that "vocabulary accounts for a large part of the variance on comprehension tests" are permissible and defensible.

It has been well established that: (1) not all comprehension test items depend on information in the passages for the correct answers (Pyrczak, 1972; Tuinman, in press), (2) substantial deletion of content words in reading passages on comprehension tests interferes only moderately with performance on the comprehension items (Weaver & Bickley, 1967), (3) the deletion of all function words in a passage results in only moderate drop in scores on a multiple-choice test based on that passage (Tuinman & Gray, 1972), and (4) syntactic structure alone conveys a large amount of semantic information (Perfetti, 1972).

This information indicates that part of the performance on a reading comprehension test conceivably could be under minimal control of knowledge of the meaning(s) of words used in the passages. Nevertheless, it is counter intuitive to construe these findings as indicating that, normally, knowledge of word meanings would not be helpful in comprehension of a passage.

The research hypothesis we set out to investigate was: Given some baseline performance on a comprehension test, typical for a specific group of subjects, the scores on that test can be raised considerably by explicit teaching of the meaning of difficult words in the tests' paragraphs, provided that these paragraphs contain a sizeable number of words, the meanings of which are not well known as measured by a recognition vocabulary test.

Procedures

Development of Materials

Comprehension tests. The reading comprehension subtest of the 1970 edition of Level 3 (for Grades 4-6) of the *California Achievement Test* was selected for the comprerension measure. The study skills items were omitted from the test, resulting in a mimeographed version of three passages and 28 items. There were two forms of the test, A and B, taken, respectively, from Forms A and B of the complete test.

The *California Achievement Test* was chosen because an earlier study (Tuinman, in press) demonstrated that the items in Form A of this test showed above-average passage dependency.

Selection of vocabulary words. The passages in each test were read and all words whose meaning appeared to be unknown or difficult for 4th, 5th and 6th graders were tentatively identified. This subjective procedure was cross-validated by looking up the difficulty values in *Children's Knowledge of Vocabulary Words* (Dale & Eicholz, 1960). This procedure resulted in two lists of words ranging in difficulty from the 6th to the 12th grades. Sixty words were selected from Form A and 66 from Form B. Sixty-three words which did not appear in either Form A or B of the comprehension tests were selected from the *Children's Knowledge of Vocabulary Words* for the Form C words. These words were approximately of the same difficulty level as the A and B words.

Vocabulary tests. Three vocabulary tests, VA, VB, and VC, were constructed to test the students' knowledge of the selected words. These tests were of the familiar multiple-choice format with the key word as the stem and four options which were either words or brief phrases. The same test was used as a pre- and posttest.

Instructional materials. The words selected for each form were divided into 11 groups of approximately six words each. Each exercise group consisted of five parts: (1) the definitions of the words, (2) sentences using the words in which the student was to determine whether the word made sense in the sentence, (3) fill-in-the-blank exercises, (4) open-ended application questions using the vocabulary words, whose meanings the student had to know to answer the questions and (5) supplementary exercises in which the student was to illustrate the words. Keys were supplied for each exercise set, so that the materials would be largely self-instructional. One final set of review material consisting of fill-in-the-blank exercises was constructed for each group of words.

Experimental Procedures

An elementary school in central Indiana was the site of the study. Nine classes, three each in grades 4, 5 and 6 with a total of 210 students, participated. A number of orientation meetings were held with the participating teachers.

Sixteen class periods of approximately 45 minutes each were needed to complete the study. The sessions extended over a period of five weeks. In the first session, the vocabulary pretests were administered. The tests for each classroom were stacked alternately so every third pupil was assigned to an A, B, or C treatment group. The students did not receive any feedback on the results of the vocabulary tests until the experiment was completed.

In sessions 2-12, students individually completed all instructional materials for the group to which they were assigned (either A, B, or C), regardless of their score on the vocabulary pretest. Students were told not to discuss the words they were studying with other students. Sessions 13 and 14 were review sessions. During sessions 2-14 one experimenter visited the school several times for observation and answering questions. All teachers seemed to be extremely cooperative in following the experimental procedures laid out in great detail in a 30-page manual of instructions and procedures.

In session 15, the vocabulary posttest was administered and in session 16 Forms A and B of the comprehension tests were administered to all students. Students were not told that there was no C comprehension test until the tests were administered.

Results

The nature of the investigation made the data from those subjects who achieved relative mastery on the vocabulary posttest of particular interest. These subjects will be designated the 80+ group, indicating a score of over 80 percent on the vocabulary posttest. In experimental group A (n=64), 33 subjects achieved the criterion score, 44 in group B (n=66) and 48 in group C (n=61). The entire original treatment group will be referred to as the "full" group.

Learning from Vocabulary Instruction

As expected, all groups improved their scores on the vocabulary test from pretest to posttest. Table 1 shows the size of the differences, an average of about 20% over the original score for the full group and approximately 23% for the 80+ group. This table also shows that for groups A and C the variance decreased considerably from the pretest to posttest. An unexplained exception is formed by the children in the B group.

The differences between pre- and post-vocabulary tests are statistically reliable as shown by a repeated measures ANOVA (F=394.43, df=1,180, p<.001).

The information of real importance with regard to the vocabulary scores is the fact that quite a few subjects reached the 80% criterion

score. For the 80+ group, the posttest mean score for group A equals 89%, for group B, 88% and for group C, 94%. For the full group these values are 74%, 78% and 85%, respectively.

Table 1
Means and Standard Deviations for the 80+ Group and the Full Group

Group	Test	80+ Group Mean	80+ Group SD	Full Group Mean	Full Group SD
A	VP	38.15	(8.50)	32.95	(9.89)
	VPO	53.45	(3.07)	44.08	(12.52)
	CA	21.09	(3.22)	17.48	(5.42)
	CB	17.57	(4.47)	14.31	(5.10)
B	VP	48.39	(13.65)	41.76	(15.43)
	VPO	59.64	(13.61)	52.27	(16.85)
	CA	19.64	(5.50)	16.73	(6.37)
	CB	17.89	(6.33)	15.27	(6.84)
C	VP	42.71	(9.61)	38.67	(12.11)
	VPO	58.94	(4.08)	53.57	(12.03)
	CA	19.63	(4.71)	18.30	(5.23)
	CB	17.83	(5.19)	15.93	(6.01)

VP = Vocabulary Pretest
VPO = Vocabulary Posttest
CA = Comprehension Test A
CB = Comprehension Test B
VP-A = 60 items
VP-B = 68 items
VP-C = 63 items

Performance on the Comprehension Tests

In order to provide some perspective on the performance of the children in this experiment, the mean score for the 28 items of CA, obtained by a sample of 523 fourth, fifth and sixth graders in ten school systems across Indiana, was calculated. The data were obtained from Tuinman (in press). This mean score was 17.49.

Again, we refer to Table 1 for the mean CA and CB scores of the three groups in the present experiment. Even casual inspection makes it clear that neither group A nor group B profited noticeably from learning the respective vocabulary. Comparison of the means on CA and CB for group A, B and C makes that quite clear. The largest difference in favor of specific word study is 1.45 (Group A, test CA, versus Group B, test CA) for the 80+ group.

More formally tested, the results of a repeated measures ANOVA on the data of the 80+ group. For the purpose of this analysis, equal cell numbers (n=33) were obtained by random deletion of B and C

subjects. The effect of the Groups factor is nonsignificant, upholding our randomization process; the effect of Tests is highly significant, affirming the different difficulty levels of the *California Achievement Tests* A and B forms as shown in the raw score to grade equivalent conversion tables in the test manual. Most telling, however, is that the interaction of Groups and Tests is nonsignificant.

The same ANOVA for the full group was also computed. Again, the interaction between Groups and Tests is nonsignificant. Other results were virtually identical to the preceding analysis.

An exploratory ANOVA with only groups A and B on the group factor was computed. The only purpose for this analysis was to show that elimination of Group C removed virtually no variance from the interaction sum of squares. This, of course, is as one would expect if the research hypothesis had been upheld, since the C group studied neither A nor B vocabulary.

No simple main effects comparisons to investigate further the nature of the interaction of Tests and Groups were calculated. The differences on the comprehension tests are, in our view, simply too small.

When the results for individual grades are analyzed, the pattern described above holds up in these analyses as well.

Relationships between Vocabulary Tests and Comprehension Tests

Table 2 contains the correlation matrices for all three groups, one set for the "full" group and one for the 80+ group. Some interesting observations need to be made here.

Understandably, the correlations between the vocabulary and com-

Table 2
Correlations Among Vocabulary and Comprehension Tests for Full and 80+ Groups

Group		Full VP	Full VPO	Full CA	Group		80+ VP	80+ VPO	80+ CA
A		VP	VPO	CA	A		VP	VPO	CA
(N=64)	VPO	.624	—	—	(N=33)	VPO	.506	—	—
	CA	.591	.673	—		CA	.247	.247	—
	CB	.747	.634	.697		CB	.685	.470	.519
B		VP	VPO	CA	B		VP	VPO	CA
(N=66)	VPO	.884	—	—	(N=44)	VPO	.854	—	—
	CA	.883	.813	—		CA	.788	.742	—
	CB	.765	.698	.782		CB	.621	.576	.762
C		VP	VPO	CA	C		VP	VPO	CA
(N=61)	VPO	.786	—	—	(N=48)	VPO	.734	—	—
	CA	.753	.661	—		CA	.672	.542	—
	CB	.795	.681	.750		CB	.740	.546	.675

prehension tests are larger for the pretests than for the posttests, particularly for the 80+ group. This is merely an artifact of ceiling effects and generally restricted variance on the vocabulary posttests as a function of learning.

Of some interest is the fact that, generally, performance on CB is more highly correlated with the VP-A and the VPO-A, whereas performance on CA is better predicted by VP-B and VPO-B. In multiple regression equations based on the correlations for the full group, 49% of the variance on CA is accounted for by VP-A and VPO-A, but these same tests account for 59% of the variance on CB. Similarly, VP-B and VPO-B account for 59% of the variance on CB, but for 77% on CA. For VP-C and VPO-C, these percentages are 64 and 58, respectively. Incidentally, inclusion of VPO into the regression equations generally accounts for but a few percentage points of the variance. The equations discussed are listed in Table 3.

Table 3

Regression Equations Predicting Comprehension Scores on the Basis of Vocabulary Scores; Full Groups (A, N=64; B, N=66; C, N=61)

1.	CA=3.33+.22 VPO-A+.14 VP-A	
	R=.70	R^2=.49
2.	CB=.07+.30 VP-A+.10 VPO-A	
	R=.77	R^2=.59
3.	CA=.79+.21 VP-B+.13 VPO-B	
	R=.84	R^2=.72
4.	CB=.65+.30 VP-B+.04 VPO-B	
	R=.77	R^2=.59
5.	CA=3.88+.26 VP-C+.08 VPO-C	
	R=.76	R^2=.58
6.	CB=−1.04+.34 VP-C+.07 VPO-C	
	R=.80	R^2=.64

Discussion

The difficulty level for naive subjects of both comprehension tests is around 60%. That leaves room for an increase in performance of 40%. How does one effect such a gain or a considerable part of it? It seems logical to turn first to the major known sources of variance on comprehension tests, vocabulary and reasoning.

We interpreted "word-knowledge" as knowledge of the specific meaning of a word as used in a passage. We demonstrated that our students' pre-experimental knowledge of these words, as measured by our multiple-choice recognition tests, was not perfect. The mean scores

on the vocabulary pretests were around 60% for the full groups and only five to ten percentage points higher for the 80+ groups. We were successful in improving these vocabulary scores considerably. Yet hardly any worthwhile effect on the comprehension scores resulted.

Given a particular comprehension item, one can generate all kinds of hypotheses about why that item does not depend on word knowledge. Specific word meanings seem to be but one usable source of information.

Word order for one, the presence of bound morphemes such as tense markers for another, can convey all kinds of information useful in answering comprehension questions. And so can other sources, the utilization of which can be called reasoning.

Yet the contribution of increased word knowledge to comprehension in this experiment was surprisingly small. If the results are reliable, they seem to rule out at least one interpretation of the statement that "word knowledge accounts for a large part of the variance on comprehension tests."

It is conceivable that the type of word knowledge measured and taught is at fault. In Fillmore's (1970) terminology, we focused largely on the basic sense characterizations of the words, whereas it is not unlikely that a teaching program emphasizing such characteristics as inherent arguments, implicit arguments and other selection restrictions would result in a kind of word knowledge that would be more instrumental in responding to comprehension test items.

Though we have decided to keep measures of reading comprehension as they are used as a given, it makes sense to consider the independent variables of interest, such as "word knowledge" and "reasoning," and to attempt a redefinition and reoperationalization of these two concepts which incorporates current thinking in psychosemantics. The question of how vocabulary accounts for variance on reading comprehension tests is extremely interesting. If we want to teach comprehension effectively, more answers to this question and the ones it entails are required.

References

Dale, E., & Eicholz, G. *Children's knowledge of words, an interim report.* Columbus: Bureau of Educational Research and Science, Ohio State University, 1960.

Davis, F. B. Psychometric research on comprehension in reading. In F. B. Davis (Ed.), *The literature of research in reading with emphasis on models.* Graduate School of Education, Rutgers — the State University of New Brunswick, 1971, section 8.

Fillmore, C. J. Types of lexical information. In F. Kiefer (Ed.), *Studies in syntax and semantics.* Dordrecht, Holland: D. Reidel, 1970.

Palermo, D. S., & Molfese, D. L. Language acquisition from age five onward. *Psychological Bulletin,* 1972, *78,* 409-428.

Perfetti, C. A. Psychosemantics: Some cognitive aspects of structural meaning. *Psychological Bulletin,* 1972, *78,* 241-259.

Pyrczak, F. Objective evaluation of the quality of multiple-choice items designed to measure comprehension of reading passages. *Reading Research Quarterly,* 1972, *8,* (1), 62-72.

Tuinman, J. J. Determining the passage dependency of comprehension questions in 5 major tests. *Reading Research Quarterly* (in press).
Tuinman, J. J., & Gray, G. The effect of reducing the redundancy of written messages by deletions of function words. *Journal of Psychology*, 1972, *82*, 299-306.
Weaver, W. W., & Bickley, A. C. Sources of information for responses to reading test items. Paper presented at the annual meeting of the American Psychological Association, Washington, September 1967.

TEST REFERENCES

Tiegs, E. W., & Clark, W. W. *Examiner's manual: California achievement tests, 1970 edition.* Monterey, California: CTB/McGraw-Hill, 1970.

THE ROLE OF VOCABULARY KNOWLEDGE IN COMPREHENSION

R. PAUL STRATTON
PHIL L. NACKE
University of Kentucky

Factor analytic studies of reading skills consistently have identified vocabulary knowledge as a major contributor to reading comprehension. Such studies have also identified, with some inconsistency, a group of highly correlated factors which account for less variance than vocabulary knowledge and include, for example, the ability to recognize the author's purpose and to follow the structure of the passage (e.g., Davis, 1968; Spearritt, 1972; Thorndike, 1971). One implication of such research for teaching reading is that a major effort should be directed toward building vocabulary skills. Recent efforts to improve reading comprehension scores on standardized tests by teaching the meanings of unfamiliar words have been unsuccessful (Tuinman and Brady, 1974). Such data indicate that a re-examination of the role of vocabulary knowledge in comprehension is in order.

A review of the research on reading comprehension strongly indicates that, while knowledge of word meanings is important, complete vocabulary knowledge, in itself, is neither a necessary nor a sufficient condition for comprehension. That is to say, reading is a complex information-processing task and identifying word meanings is but one element in the total process. The mature reader has acquired an adequate variety of operations which permit comprehension even though the meanings of all of the words in the discourse may not be available (necessity). Moreover, these additional operations are of such importance in the total process that failure of one operation may preclude the utilization of word meanings even though they may be available (sufficiency).

The purpose of this paper is to summarize some of the research relating to the role of vocabulary knowledge in comprehension and to discuss this issue within a levels of processing framework.

Necessity

Characteristics of the English language which support the notion that complete vocabulary knowledge is not necessary for comprehension are its redundancy and predictability. Supportive of this point of view is the competent language user's ability to expand ideas and to handle language in which words have been deleted, such as in telegrams, newspaper headlines, titles, and tests employing the cloze procedure.

It is apparent that not all words in prose are of the same information value. Consider the following paragraph in which the syntactically