DOCUMENT RESUME

ED 162 277   CS 004 481

AUTHOR   Johnson, Dale D.; And Others
TITLE   On-Line Diagnosis of Reading Difficulties-IV.
                Technical Report No. 464.
INSTITUTION   Wisconsin Univ., Madison. Research and Development
                Center for Individualized Schooling.
SPONS AGENCY   National Inst. of Education (DHEW), Washington,
                D.C.
PUB DATE   Sep 78
CONTRACT   NIE-G-78-0217
NOTE   60p.

EDRS PRICE   MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS   Cloze Procedure; *Computer Assisted Instruction;
                Context Clues; Educational Diagnosis; Elementary
                Education; Reading Comprehension; Reading Diagnosis;
                *Reading Difficulty; *Reading Skills; *Test Items;
                Tests; *Vocabulary Skills
IDENTIFIERS   PLATO Elementary Reading Curriculum Project.

ABSTRACT

            To evaluate the performance of elementary school
children on various paradigms for assessing vocabulary knowledge
on-line, three studies were conducted. In the first, 173 second
through eighth grade students were involved in an investigation of
five approaches to vocabulary assessment: synonym in context, synonym
out of context, cloze, oral recognition, and self-screening. Results
revealed significant problems with the self-screening and oral
recognition formats; they were therefore eliminated from the
succeeding studies. A total of 184 first through fifth graders
participated in the second and third studies, which further evaluated
the three remaining formats. These studies also correlated the
children's performance on the three paradigms with scores on a
standardized measure of reading comprehension. The results suggest
that no single format is superior in assessing vocabulary knowledge
across all grade levels. Data from the third study also indicate
lower correlations between vocabulary knowledge and comprehension
ability for the second and third grades than for the fourth and fifth
grades. (Author/FL)

Technical Report No. 464

ON-LINE DIAGNOSIS OF READING DIFFICULTIES-IV

by

Dale D. Johnson, Susan D. Pittelman,
Judy Schwenker, and James Perry

Report from the Project on
Studies in Language: Reading and Communication

Dale D. Johnson
Faculty Associate

Wisconsin Research and Development
Center for Individualized Schooling
The University of Wisconsin
Madison, Wisconsin

September 1978

The project presented or reported herein was performed pursuant to a grant from the
National Institute of Education, Department of Health, Education, and Welfare. However,
the opinions expressed herein do not necessarily reflect the position or policy of the
National Institute of Education, and no official endorsement by the National Institute
of Education should be inferred.

11

3

MISSION STATEMENT

The mission of the Wisconsin Research and Development Center
is to improve the quality of education by addressing the full
range of issues and problems related to individualized schooling.
Teaching, learning, and the problems of individualization are
given concurrent attention in the Center's efforts to discover
processes and develop strategies and materials for use in the
schools.   The Center pursues its mission by

● conducting and synthesizing research to clarify the
  processes of school-age children's learning and
  development

● conducting and synthesizing research to clarify effective
  approaches to teaching students basic skills and concepts

● developing and demonstrating improved instructional strategies,
  processes, and materials for students, teachers, and school
  administrators

● providing assistance to educators which helps transfer the
  outcomes of research and development to improved practice
  in local schools and teacher education institutions

The Wisconsin Research and Development Center is supported
with funds from the National Institute of Education and the
University of Wisconsin.

                                WISCONSIN RESEARCH AND DEVELOPMENT
                                CENTER FOR INDIVIDUALIZED SCHOOLING

Acknowledgments

The work reported here was done in cooperation with the PLATO

Reading Project at the University of Illinois. The assistance

of Priscilla Obertina is gratefully acknowledged.

We also acknowledge with gratitude the assistance of the

teachers and pupils at Reinberg School in Chicago, Illinois and

at Columbia School in Champaign, Illinois; Dr. Hazel Wagner of

the Department of Research and Evaluation of the Chicago Public

Schools; Dr. Chris Dimas, Director of the PLATO Project for the

City Colleges of Chicago; and Dr. Richard Neapolitan, PLATO

Coordinator at Wright Community College, Chicago, Illinois.

5

Table of Contents

6

v

Table of Contents (Con't.)

List of Tables

8

List of Tables (Con't.)

List of Figures

10

On-Line Diagnosis of·Reading Difficulties-IV

This is the final report in a series of four on the feasibility of using the PLATO system for on-line diagnosis of reading skills. In the first study (Venezky, Bernard, Chicone, & Leslie, 1975), a comparison was made between on-line and paper-and-pencil administration of a comprehension test for fifth and sixth graders. This study demonstrated that equivalent scores could be obtained and that PLATO was usable at these grade levels for comprehension testing.

A second study (Venezky, Chicone, & Perry, 1976) explored the feasibility of on-line testing of letter-sound encoding with first, second, and third graders. The results of these studies were mixed. While five of the nine patterns tested showed results similar to those obtained by standard testing procedures, limitations were found in the quality of the PLATO audio recordings and in the testing paradigm which was selected.

The third study (Venezky, Perry, Chicone, & Pittelman, 1977) examined three on-line techniques for assessing vocabulary knowledge: self-screening, word matching under brief exposure, and synonyms. At the same time, the study provided some preliminary information on response-contingent testing. Brief descriptions of all tests developed, as well as an on-line bibliography of reading research literature, were given in the Appendices of the report documenting

that study.

This final set of feasibility studies is a continuation and refinement of the work begun in the third study. This current study compares the performance of elementary school children of various reading abilities on five different paradigms for assessing vocabulary knowledge: synonym in context, synonym out-of-context, cloze, oral recognition, and self-screening. All development work was done at the Wisconsin Research and Development Center on a PLATO terminal which included a touch detection collar and a printer. Testing for the first two studies discussed in this paper was conducted at Wright Community College in Chicago, Illinois. Testing for the third study was conducted at Columbia Elementary School in Champaign, Illinois.

STUDY I

Study 1 examined the characteristics of five different approaches to vocabulary testing: self-screening, oral recognition, synonym out-of-context, synonym in context, and cloze. All formats, with the exception of oral recognition, were administered on-line using the PLATO terminal. Two formats, self-screening and synonym out-of-context, had been tested in an earlier study (Venezky et al., 1977), but many of the stimulus words and distractors used were revised.

The Self-screening Test required subjects simply to indicate

whether or not they knew the meaning of a target word. The Oral
Recognition Test, a paper-and-pencil test, required subjects to
find the exact match to a spoken word from among four alternatives.
The Synonym out-of-context Test required subjects to choose, from
among four alternatives, the word closest in meaning to a target
word. The Synonym in Context Test presented the target word within
the context of a sentence, and then required subjects to choose, from
among four alternatives, the word closest in meaning to the target
word. The Cloze Test required subjects to select, from among four
alternatives, the word which best fit in a sentence that had one word
deleted.

Method

Each of the five test formats had seven levels of difficulty
with ten items per level. (In the previous study [Venezky et al.,
1977] the Self-screening Test had six different levels with eight
items per level.) The same stimulus words were used on all five
test formats, so that a direct comparison of test paradigms could
be made. Words were selected from several standardized vocabulary
tests at each level to represent typical reading vocabularies from
grades one through seven. Each test level represented a single
grade level of difficulty. The stimuli for Level 4 of the five
test paradigms are shown in Appendix A.

Subjects

A total of 173 elementary school students in grades 2 through
8 of an inner city Chicago elementary school participated in the
study. Testing was done in two separate sessions. During the first

session, which took place in December 1976, 107 children were tested. The remaining 66 children were tested during the second session, in January 1977. Each testing session lasted for two days.

Procedure

The subjects were brought from their classrooms in groups of 24 to the testing site at Wright Junior Community College. The college was located approximately three blocks from the school. Each testing period began with a short orientation, in which the subjects were assigned a "PLATO name" (their own first name and any initial letters of their last name required to resolve same first name conflicts). Subjects were then shown two displays in order to familiarize them with the equipment and with the procedures for taking the tests. During these practice displays, subjects were instructed as to how to select and change an answer.

The sign-on procedures took approximately three to five minutes, with the longer times taken by younger subjects. Individual help, with signing-on was given to all second and third grade subjects, because these younger pupils had difficulty locating specific letter keys on the standard typewriter keyboard. Older subjects were given help only if they requested it. The practice displays took from three to five minutes to complete, with younger subjects requiring the longer times.

Each subject was given the Self-screening Test, the Oral Recognition Test, and either the Synonym out-of-context Test, the Synonym

in Context Text, or the Cloze Test. During the first testing

session (December 1976), subjects began the test session with

the Self-screening Test, followed by either the Synonym out-of-

context Test, the Synonym in Context Test, or the Cloze Test.

Subjects were then taken from the PLATO terminals to tables for

the paper-and-pencil Oral Recognition Test.

During the December testing, examiners observed that a large

number of subjects responded correctly to items on the Oral

Recognition Test before the stimulus word was spoken. (Subjects

apparently circled words which they remembered seeing moments

before on the two PLATO tests.) This effect was especially appar-

ent among older subjects. As a result, the order of administra-

tion of the tests was changed before the second testing session

(January 1977). In the January testing session, subjects were

first given the Self-screening Test, followed by the Oral Recogni-

tion Test, and then either the Synonym out-of-context Test, the

Synonym in Context Test, or the Cloze Test. The change in test

order required a small programming adjustment which locked subjects

on what had been a "resting" page. Subjects were then directed to

tables where the Oral Recognition Test was administered. After

completion of the Oral Recognition Test, subjects returned to their

terminals and were told which keys to press in order to view the

last test.

Self-screening Test. Figure 1 illustrates the layout of a

PLATO page from the Self-screening Test. This test was the first

one administered. The subject was presented with five words at

a time and was to touch those words whose meaning he or she did

not know. When a word was touched, a box appeared around it.

Touch the words whose meaning you DON'T know.

hesitate   menacing

                        confident

gigantic   briskly

If you know the meaning of all the words, press NEXT
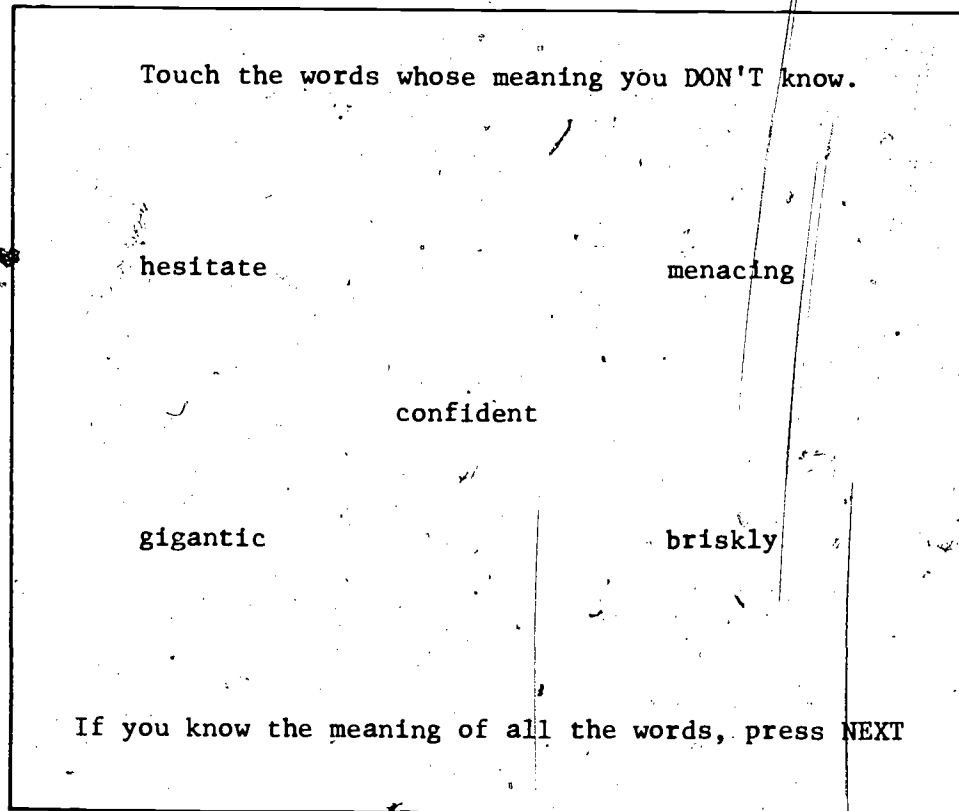
Figure 1. The layout of a PLATO page from the

Self-screening Test.

A second touch removed a box mistakenly placed around a word.

After each set of five words had been inspected and the appropriate

words touched, the subject pressed the NEXT key on the keyboard

to see the next set of words. After the Self-screening Test was

completed, the subject was shown a short diversionary display and

invited to rest for a few seconds. In the January testing, subjects were then directed to tables for the Oral Recognition Test.

Oral Recognition Test. The Oral Recognition Test was a paper-and-pencil test. The test administrator pronounced a target word and subjects were to circle that word from a set of four response words. The four response words for each item were arranged in a horizontal row to the right of the number for that item (see Figure 2 on following page). After the Oral Recognition Test, subjects returned to the terminals and pressed the NEXT key, which led them to one of the three tests described below.

Synonym out-of-context Test. Subjects were shown a word in a box near the top of the screen and four response words below it (see Figure 3).

Touch the word which is closest in meaning

to the word in the box

briskly

awkwardly   sadly   rapidly   slowly

If you are finished, press NEXT

OR, to change your answer touch another word.

Figure 3. The layout of a PLATO page from

the Synonym out-of-context Test.

Name _____   ORAL RECOGNITION TEST--4

Grade _____

1.	briskly	biscuit	blackie	briefly
2.	freedom	triple	feeble	table
3.	state	estate	hasty	hesitate
4.	gentle	gigantic	gyration	gymnastic
5.	concealed	concrete	concubine	convinced
6.	powder	paucity	paper	pauper
7.	confetti	convince	confident	condemn
8.	mention	menacing	marking	mendacity
9.	obvious	abound	obtain	abstain
10.	purpose	persist	purse	pursued

Figure 2. An example of a page from the paper-and-pencil

administered Oral Recognition Test.

18

The subject was to touch the word which was closest in meaning to the word in the box. When a response word was touched, a box appeared around it. If the subject wanted to change the answer, he or she could do so by touching another response word. The erroneous box would erase and the newly selected word would receive a box.

Synonym in Context Test. Subjects were shown a sentence in which one of the words was underlined, as in Figure 4. Below the sentence were four response words. The subject was to touch the word which was closest in meaning to the underlined word in the sentence.

Touch the word which is closest in m ing to
the underlined word.

The feeble old dog walked beside the woman.

funny   weak   strong   fleet

If you are finished, press NEXT.

OR, to change your answer touch another word.

Figure 4. The layout of a PLATO page from
the Synonym in Context Test.

When a word was touched, a box appeared around it. If the subject wanted to change the answer, he or she could do so by touching another response word. The erroneous box would erase and the newly selected word would receive a box.

Cloze Test. Subjects were shown a sentence in which there was a blank space in place of one of the words (see Figure 5).

```
┌──────────────────────────────────────────────────────────┐
│                                                            │
│     Touch the word which best completes the sentence       │
│                                                            │
│                                                            │
│     The horse trotted_____along a familiar path.          │
│                                                            │
│                                                            │
│     after         below         twelve         bris        │
│                                                            │
│                                                            │
│            If you are finished, press NEXT.                │
│                                                            │
│         OR, to change your answer touch another word.       │
│                                                            │
└──────────────────────────────────────────────────────────┘
```

Figure 5. The layout of a PLATO page from the Cloze Test.

Below the sentence, were four response words, one of which would make the sentence complete. The subject had to touch the response word which would best complete the sentence. When a word was touched, a box appeared around it. If the subject wanted to change the answer, he or she could do so by touching another response word. The erroneous box would erase and the newly selected word would receive a box.

Each subject took three levels of difficulty for each test paradigm assigned: a test one grade below grade level, a test at grade level, and a test one grade above grade level. The three levels were presented in order from below grade level, to at grade level, to above grade level. (Sixth, seventh, and eighth grade subjects received levels five, six, and seven, because level seven was the highest level that had been developed.) Total completion time for all three levels of the three formats was between 20 and 30 minutes. Average testing times were nearly the same for both testing sessions, even through subjects in the January session were required to move away from and then back to the PLATO terminals.

Results

Table 1 presents mean percentage correct scores listed by test format and by test level.

Table 1

Mean Scores (% Correct) by Test Level and Format

Level	Self-screening	Synonym out-of-context	Cloze	Synonym in context	Oral recognition
1	95	45	87	29	95
2	73	55	69	55	92
3	60	47	50	51	92
4	43	42	58	56	95
5	63	68	79	75	97
6	72	57	88	80	100
7	71	67	88	86	100

Self-screening

The results of the self-screening format, arranged by test level and grade, are shown in Table 2. Figure 6 shows these same results in mean scores for each grade collapsed across test levels.

The most notable aspect of these results is the higher scores of the primary (first, second, and third) and upper (seventh and eighth) grades as compared to the middle grades (fourth, fifth, and sixth). The higher scores of children in the primary grades may be due to three factors: (a) During testing, it was observed that many of the younger subjects did not always use the PLATO keyboard appropriately. As a result, some subjects hit the "NEXT" key twice in succession. This caused a display of five words to be recorded as words that were known by subjects, even though the five words had appeared only briefly on the visual display. (b) Younger subjects may have been overconfident of their knowledge of the word definitions. This is indicated by their relatively lower scores on the other test formats. (c) Word items for these lower grade levels were relatively more common than words appearing in the middle test levels.

The higher scores of pupils in the upper grades (seventh and eighth) can be partially accounted for by the testing design. Seventh and eighth grade subjects took the same three levels of the test as did sixth graders.

Some of the confounding factors discussed above could be eliminated by altering the PLATO program to prevent false positive marking of vocabulary words and by changing and extending the word list. It

Table 2

Mean Scores (%Correct), Standard Deviation, and Number of

Subjects on Self-screening Format by Test Level and Grade

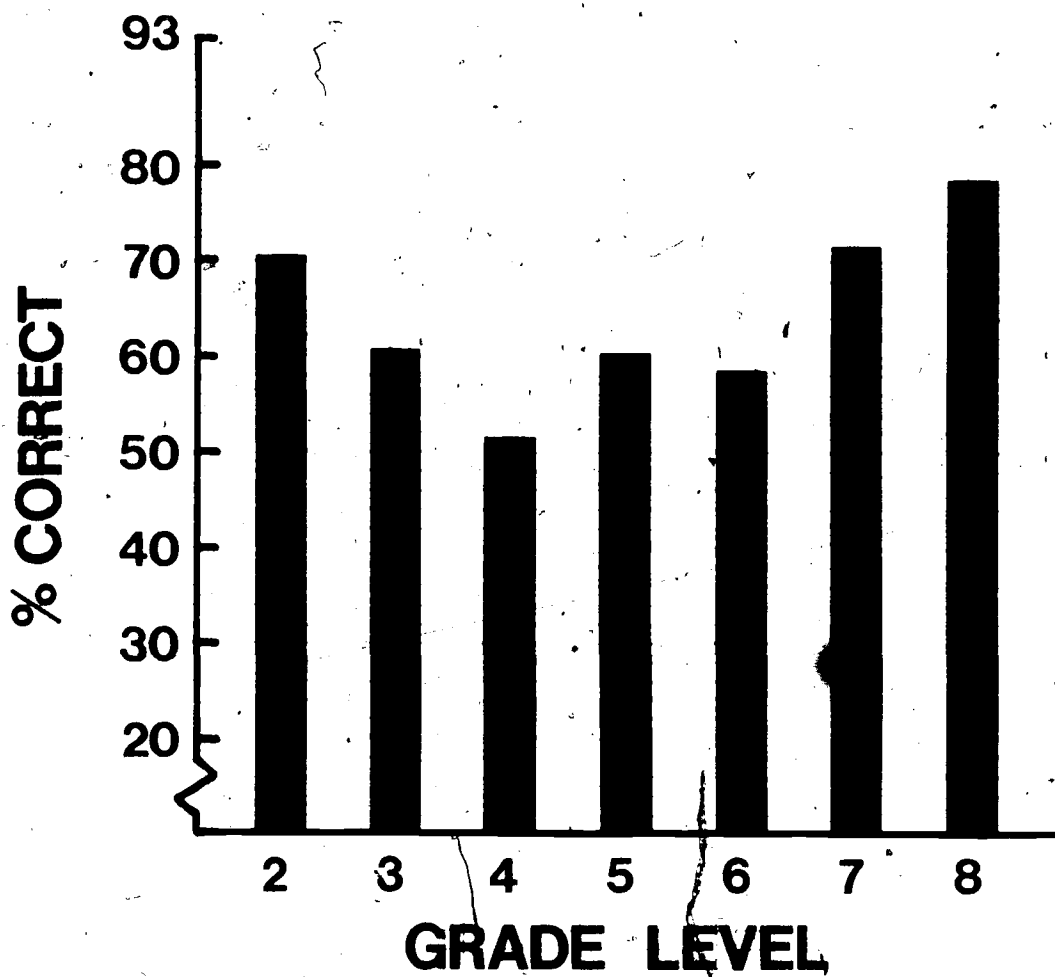| Test level	Grade						
	2	3	4	5	6	7	8
1							
$\overline{X}$	95						
SD	13						
N	22						
2							
$\overline{X}$	68	77					
SD	24	21					
N	22	22					
3							
$\overline{X}$	54	60	69				
SD	30	29	24				
N	22	22	23				
4							
$\overline{X}$		45	39	49			
SD		27	24	22			
N		22	23	13			
5							
$\overline{X}$			47	58	53	69	80
SD			22	22	17	17	11
N			23	13	25	34	34
6							
$\overline{X}$				74	61	74	79
SD				17	14	14	11
N				12	25	34	34
7							
$\overline{X}$					62	72	77
SD					14	15	13
N					25	34	34

Figure 6. Mean scores (% correct) on

self-screening format, by grade.

was decided, however, that the "overconfidence" effect would still be present, and would cause difficulty in the interpretation of results. For this reason, further use of the self-selection format was not planned.

Oral Recognition

The mean percentages correct on the oral recognition format are listed in Table 3. It is apparent from the data in Tables 1 and 3 that there is a pronounced ceiling effect for the Oral Recognition Test. This suggests that the test was too easy for the grades tested.

Since the target words were the same as that on the other test formats, the task was apparently not measuring the vocabulary skills which were of interest to the study. For this reason, no further consideration was given to the oral recognition format.

Synonym out-of-context

Table 4 lists the mean scores for the synonym out-of-context format shown by test level and by grade. Figure 7 plots the mean scores for each grade. The graph shows a linear increase with a noticeable dip in the fourth grade average. Table 4 indicates that level four test scores were extremely low. Item analysis for this test revealed that of the 10 items at that level, as many as seven (70%) could have been two difficult for the grades tested. When a new reference source later became available, it was discovered that four of the 10 target words in level four were considered to be seventh to ninth grade words (Dale & O'Rourke, 1976). Two other

Table 3

Mean Scores (% Correct) on

Oral Recognition Format by Grade and Test Level

Test Level	Grade						
	2	3	4	5	6	7	8
1							
$\overline{X}$	95						
SD	10						
N	22						
2							
$\overline{X}$	87	97					
SD	15	6					
N	22	22					
3							
$\overline{X}$	89	95	93				
SD	15	13	12				
N	22	22	23				
4							
$\overline{X}$		95	94	100			
SD		12	15	0			
N		22	23	13			
5							
$\overline{X}$			90	94	93	99	99
SD			15	10	4	4	3
N			23	13	25	34	34
6							
$\overline{X}$				100	100	99	100
SD				0	0	2	2
N				12	25	34	34
7							
$\overline{X}$					100	100	100
SD					0	0	2
N					25	34	34

Table 4

Mean Scores (% Correct) on Synonym out-of-context

Format by Grade and Test Level

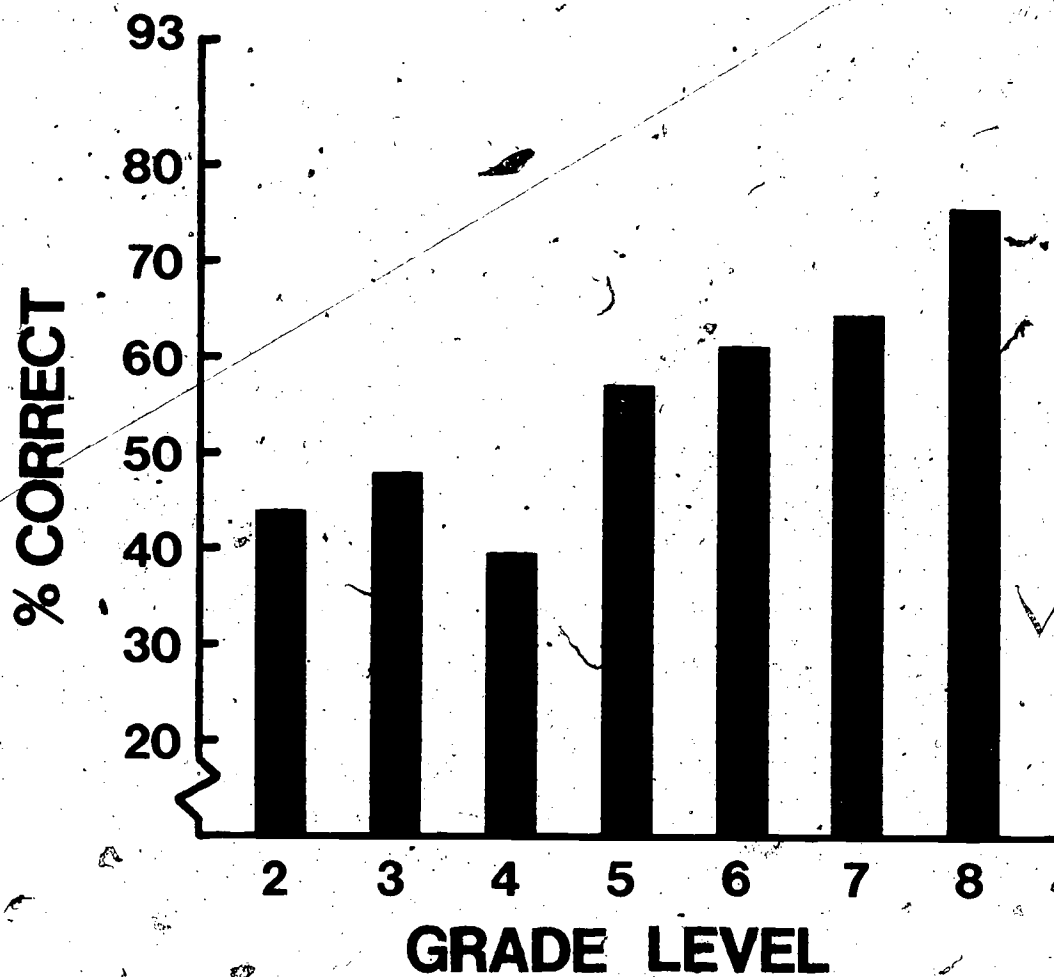| Test level	Grade						
	2	3	4	5	6	7	8
1							
$\overline{X}$	45						
SD	30						
N	8						
2							
$\overline{X}$	51	60					
SD	23	22					
N	8	7					
3							
$\overline{X}$	36	47	59				
SD	23	30	11				
N	8	7	7				
4							
$\overline{X}$		37	29	73			
SD		31	11	22			
N		7	7	4			
5							
$\overline{X}$			31	53	68	73	88
SD			15	29	15	22	13
N			7	4	6	12	12
6							
$\overline{X}$				45	53	55	67
SD				10	18	21	17
N				4	6	12	12
7							
$\overline{X}$					63	65	72
SD					14	21	14
N					6	12	12

Figure 7. Mean scores (% correct) on synonym
out-of-context format, by grade.

target words were listed as fifth to seventh grade words. This factor
alone could account for the relatively low performance by subjects
on level four. This experience also highlights the inadvisability of
using standardized vocabulary tests as sources of grade-appropriate
vocabulary items. Item 8 from Level 4 of the Synonym out-of-context
Test (shown below) illustrates the difficulty described above. The
figure in parenthesis is the percentage of the national sample of
students (at the grade level indicated) reported as knowing the words
(Dale & O'Rourke, 1976).

Target word: "menacing"   (70% at eighth grade)

Distractors: "dangerous"   (89% at fourth grade)

                      "inspiring"   (85% at eighth grade)

                      "regular"   (74% at fourth grade)

                      "slinking"   (72% at eighth grade)

Item analysis of Item 8 showed that 40% of the subjects chose the
correct response ("dangerous") while the remaining 60% were evenly
spread among the three other distractors. With this item as with
at least six additional items in Level 4, the response patterns appeared
to be random.

The overall rise in scores shown on the graph in Figure 7 may be
partially explained by the tendency for target words to become rela-
tively easier for subjects as grade level increases. Here too, seventh
and eighth grade subjects' scores were artificially inflated by the
topping out effect which resulted from giving these upper grade subjects

the same test levels as sixth grade subjects.

Synonym in Context

Table 5 lists the mean scores for the synonym in context format
by test level and grade. A graph of mean scores is shown in Figure
8. The overall rise in mean scores across grades strongly suggests
an increasing ability to use context in demonstrating word knowledge.
The most noticeable feature of Figure 8, however, is the exceptionally
low mean score for second grade subjects. Explanation for low scores
at this level may be that the limited second grade vocabulary prohi-
bits using sentences of any great length or complexity. The limited
vocabulary and the restriction of providing only one sentence of con-
text were considerable constraints in creating second grade level test
items. As a result, the efficacy of using synonym in context as a
diagnostic test format in the primary grades is questionable. Perhaps
with long passages to provide more contextual clues, the abilities of
primary grade subjects to use context to demonstrate word knowledge
could be more fully explored.

Cloze

Table 6 lists the mean scores by level and grade for the cloze
format. Figure 9 presents a graph of mean scores for each grade. As
does Figure 7, Figure 9 also shows an overall rise in mean scores
except for a drop at third grade level. Examination of Table 6
shows scores for test levels three and four to be very low. Item
analyses for those test levels suggest that, once again, a number
of target words, as well as a few of the distractors, may have been
too difficult for the grades assigned to them. Item 5 from Level 3

Table 5

Mean Scores (% Correct) on Synonym

in Context Format by Grade and Test Level

Test level	Grade						
	2	3	4	5	6	7	8
1 $\overline{X}$	29						
SD	20						
N	8						
2 $\overline{X}$	45	66					
SD	27	28					
N	8	7					
3 $\overline{X}$	30	61	63				
SD	11	30	29				
N	8	7	8				
4 $\overline{X}$		44	58	77			
SD		30	28	40			
N		7	8	3			
5 $\overline{X}$			43	63	84	83	88
SD			24	12	9	16	13
N			8	3	9	10	11
6 $\overline{X}$				53	80	78	89
SD				40	13	15	14
N				3	9	10	11
7 $\overline{X}$					88	86	85
SD					10	8	11
N					9	10	11

Figure 8. Mean scores (% correct) on synonym

in context format, by grade.

Table 6

Mean Scores (% Correct) on

Cloze Format by Grade and Test Level

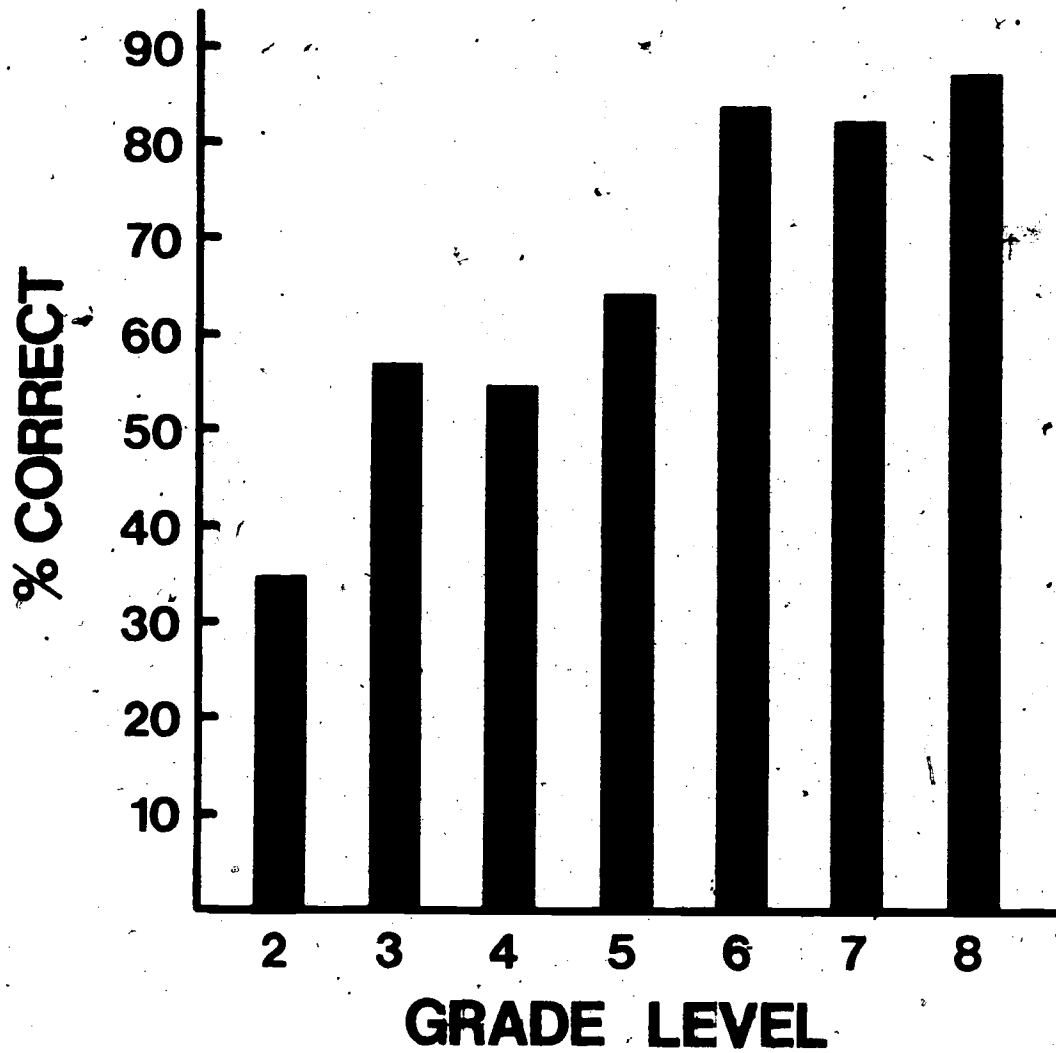| Test level	Grade						
	2	3	4	5	6	7	8
1							
$\overline{X}$	87						
SD	15						
N	6						
2							
$\overline{X}$	58	76					
SD	19	24					
N	6	8					
3							
$\overline{X}$	40	35	71				
SD	23	28	28				
N	6	8	8				
4							
$\overline{X}$		30	73	75			
SD		28	34	10			
N		8	8	6			
5							
$\overline{X}$			55	62	85	89	91
SD			26	22	11	12	10
N			8	6	10	12	11
6							
$\overline{X}$				72	89	91	90
SD				16	13	11	15
N				5	10	12	11
7							
$\overline{X}$					80	93	91
SD					12	11	17
N					10	12	11
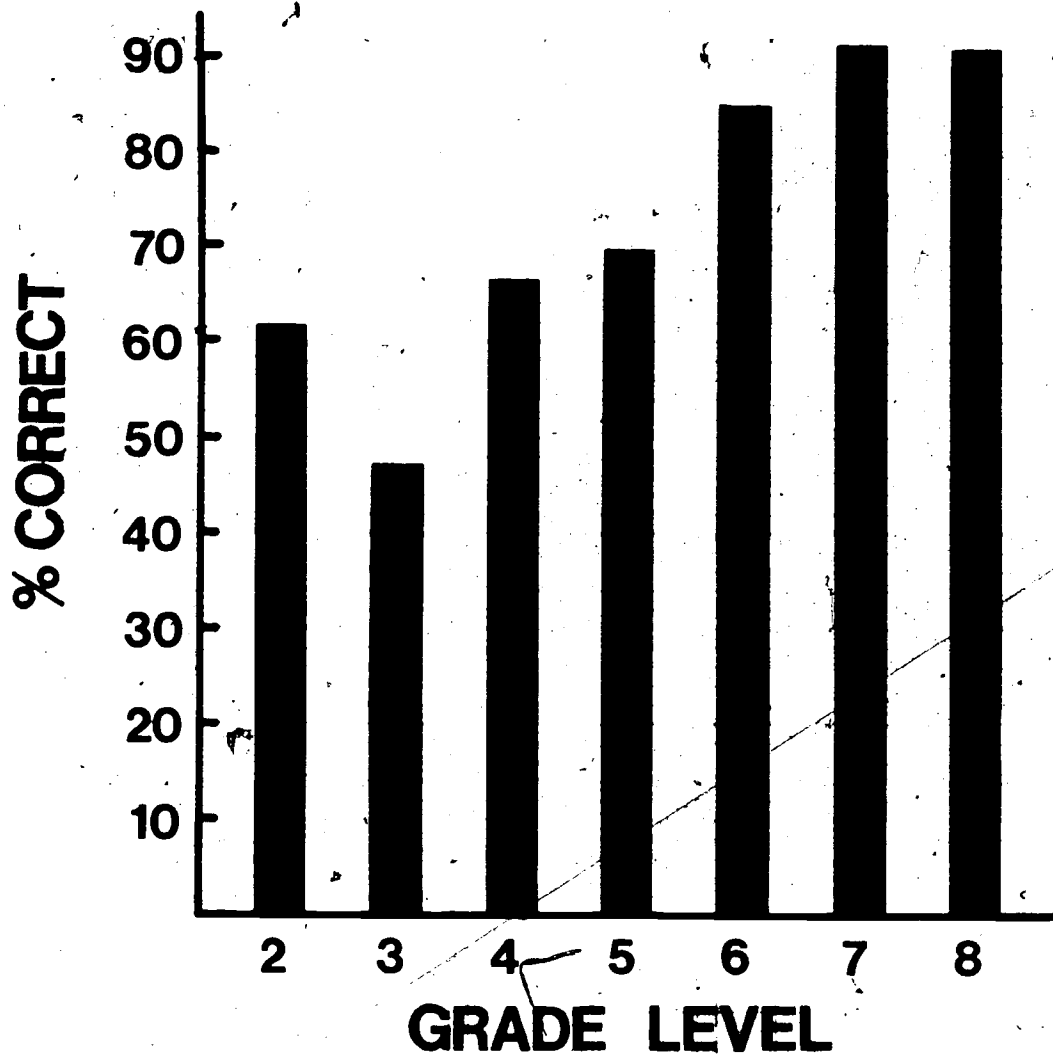
Figure 9. Mean scores (% correct) on cloze

format, by grade.

of the Cloze Test is an example:

The blue water was _____.

glistening	(82% at sixth grade)
depressing	(69% at sixth grade)
expensive	(68% at fourth grade)
cozy	(77% at fourth grade)

The figures in parentheses indicate the percentage of school children

at the grade level indicated in a nationwide sample knowing these

words (Dale & O'Rourke, 1976). In addition, there may be a problem

associated with the cloze format itself, because the context of a

brief sentence may not rule out all but the targeted response. This

presents a considerable problem when using the cloze format for

testing vocabulary in the primary grades. Corollary to this, a final

explanation for the depressed scores on Levels 3 and 4 is the novelty

of the cloze format itself. Typically, students receive little prac-

tice on cloze tasks until the later elementary grades. Finally, a

ceiling effect was noticeable in the upper grades (seventh and eighth).

That is, given the relatively easier range of difficulty for the target

words on test levels 5, 6, and 7, the scores do not reflect the same

degree of spread as might be seen if the test included levels for

eighth and ninth grades as well.

Summary

As a result of Study 1, several kinds of problems became apparent.

Some of these problems concerned the use of on-line systems for

diagnosis; others involved the testing instruments used in the study.
In particular, the false positive marking of words as known word in
the Self-screening Test (when the "NEXT" key was hit twice in succes-
sion) is a problem which can be rectified at the technical level. The
hardware must be altered to fit the situation when younger children,
unfamiliar with interactive terminals, will be using keyboards.

Other problem areas lie within the tests themselves. The feasi-
bility of drawing vocabulary items from existent standardized tests
was called into question. Consequently, more careful attention must
be given to other sources of items when these tests are revised.
Also, the two formats which utilized the context of a single sentence
(cloze and synonym in context) were discovered to be inappropriate
for the primary grade subjects. Because the ability to use context
to demonstrate word knowledge is an important vocabulary skill to
assess, alternative formats should be explored. One likely alterna-
tive would be to utilize passages, rather than single sentences, in
order to provide more contextual information within the constraints
of short sentence length and limited vocabulary usage. Finally, the
ceiling effect discernable at the upper elementary grades (seventh
and eighth) can be eliminated by the development of additional tests
for levels 8 and 9.

STUDY 2

Study 2 examined three paradigms for assessing vocabulary know-

ledge: synonym out-of-context, synonym in context, and cloze. The goal of the study was to determine which of the vocabulary formats correlated most highly with a standardized measure of reading comprehension. The three vocabulary tests were administered on-line, using the PLATO terminal, while the global reading comprehension test was a paper-and-pencil test.

The Synonym out-of-context Test required subjects to choose, from among four alternatives, the word closest in meaning to a target word. The Synonym in Context Test presented the target word within the context of a sentence, and then required the subject to choose, from among four alternatives, the word closest in meaning to the target word. The Cloze Test required subjects to select, from among four alternatives, the word which best fit in a sentence that had a word deleted. The global reading comprehension test administered to subjects was at appropriate grade levels as determined by the reading subtest of the Metropolitan Achievement Tests.

Method

The stimuli for the three vocabulary tests (Synonym out-of-context, Synonym in Context, and Cloze) were the same as in Study 1. Based on the results of Study 1, the Self-screening Test and Oral Recognition Test paradigms were omitted from further investigation. (Appendix A provides a sample of the test items that were administered.) The reading subtest of the Metropolitan Achievement Tests was administered from the test booklet. Grade one received the Primary I level, grade three received the Elementary level, and grade

28

five received the Intermediate level. All three grades received
Form F of the test.

Subjects

A total of fifty-six students from grades one, three, and five
of an elementary school in inner city Chicago, Illinois participated
in the study. The students were from the same school as in Study
1. It was anticipated that all testing would be completed in a
single school day. Because of difficulty with the main computer at
Champaign, the testing session took one and a half days. Testing
was conducted on June 2, 1977 and during the morning of June 3rd.

Procedure

Subjects were brought from their classrooms in groups of 20
to 24 to Wright Junior Community College. Subjects were
seated at tables and given directions for the reading subtest of
the Metropolitan Achievement Tests (as described in the _Teacher's
Directions Manual_ for the Metropolitan Achievement Tests). The
appropriate level of the Metropolitan Achievement Tests was then
administered to the subjects. Test administration took between 30
and 35 minutes, depending on the level of the test.

Following administration of the reading subtest of the Metropolitan
Achievement Tests, each subject was assigned a "PLATO name." The
students were then taken to the PLATO terminals and given instruc-
tions on sign-on procedures. Sign-on procedures took between three
and seven minutes, with younger students taking longer.

After the sign-on was completed, subjects were shown an inter-

active display to acquaint them with procedures for selecting and changing an answer. After the subjects were allowed sufficient time to practice these skills (approximately three to five minutes), they were given one of the three vocabulary test formats. Three levels of tests were given within a format: one test just below grade level, one test at grade level, and one test just above grade level. The tests were presented in order of level of difficulty from easiet to most difficult.

Total time at the PLATO terminal for each group of 20 to 24 students was 35 to 40 minutes.

Results

The mean scores for grades 1, 3, and 5 on the vocabulary formats are summarized in Table 7. The number of subjects tested was less than anticipated, due to difficulties with the PLATO system.

Table 7

Mean Scores (by Grade) on Three
Vocabulary Formats

Grade Level	Formats		
	Synonym in context	Synonym-out-of-context	Cloze
1	45.05 N = 8	36.28 N = 9	50.40 N = 8
3	67.14 N = 7	68.0 N = 5	68.57 N = 7
5	66.7 N = 1	48.66 N = 5	74.41 N = 6

Because of the resulting small number of subjects in each cell,

readers should be cautioned against generalizing from these data.

Despite the small number of subjects, the data in Table 8 are

interesting and infer a relationship between vocabulary knowledge

and comprehension ability. Examination of the relative size of

correlations in successive grades in Table 8, reveals that the older

the subjects, the higher the correlation between vocabulary and

comprehension scores. This was true even though item analysis in-

dicated that portions of the tests (especially at grade 3) may have

utilized target words too difficult for the designated grade level.

Table 8

Pearson Correlations of Vocabulary

Formats with Metropolitan, by Grade

| Grade Level | Formats | | |
	Synonym in context	Synonym-out-of-context	Cloze
1	.4247	.7626	.8056
3	.7420	.9712	.8335
5	----[a]	.9938	.9837

[a]Cell size too small to compute meaningful correlation.

With or without the provision of context, there was an overall

rise in mean scores, in ascending grades as well as higher correla-

tions between the vocabulary test scores and Metropolitan reading

test scores as grade level increased.

Within grades, results were less clear. The Synonym out-of-context and Cloze Tests showed good overall correlation with comprehension scores equal for all grades. The results of the Cloze Test are not surprising, as cloze is often used as a measure of comprehension. Results of using the cloze format are particularly interesting at the first grade level because children typically receive little practice on cloze tasks until the middle elementary grades. Moreover, analyses of mean scores for first grade subjects on the three vocabulary formats revealed that performance increased with the amount of contextual information provided.

Although these results should be viewed cautiously, the test data nevertheless provide valuable clues as to how elementary school children utilize context to demonstrate vocabulary knowledge. The data also indicate that no single method of vocabulary assessment correlates highly with comprehension ability across all elementary grades, thus suggesting that no one vocabulary test format is superior for all grades. The study did delineate some interesting issues for further research in the area of format effects upon vocabulary assessment performance. These issues will be examined more atomistically in future studies, through the use of group-administered paper-and-pencil tests.

STUDY 3

Study 3 was an extension of Study 2. As in the previous study,

the main goal of Study 3 was to determine which of three vocabu-
lary test formats, synonym out-of-context, synonym in context, or
cloze, correlated most highly with a standardized measure of
reading comprehension. The three vocabulary tests were adminis-
tered on-line, using the PLATO terminal, while the global reading
comprehension test was a paper-and-pencil test.

Method

The formats for the vocabulary tests, synonym out-of-context,
synonym in context, and cloze, were the same as those used in
Studies 1 and 2. Some test items were revised, however, based on
an item analysis done as part of Study 2. The stimuli for level
4 of the three test formats are shown in Appendix B.

As in Study 2, the reading subtest of the Metropolitan Achieve-
ment Tests (Form F) was used as the global measure of reading
comprehension. The test was administered by school personnel in
September 1977, and the data made available to the Wisconsin
Research and Development Center for Individualized Schooling.

Subjects

A total of 128 students from grades 2 through 5 of an elemen-
tary school in Champaign, Illinois participated in the study. The
school is located in a lower socioeconomic neighborhood. The school
uses PLATO as a means of instruction, so the pupils were already
familiar with using the terminal. Testing was conducted on November
21 and 22, 1977.

Procedure

Subjects were brought from their classrooms in groups of 10 to

the PLATO laboratory in the elementary school. Subjects were
assigned special "PLATO names" for the study, and were given a
brief practice session on how to select and change answers. Because
these students were already "PLATO-sophisticated," it was not
necessary to give extensive directions on signing-on and using
the terminals.

After a few minutes of practice, subjects were given one of
the three vocabulary test formats. As in the previous study, each
pupil was given three levels of tests within a single format. The
tests were presented in graduated levels of difficulty starting
with the one just below grade level, to the test at grade level,
and finishing with the test just above grade level. Total time
at the PLATO terminal for each group of subjects was approximately
20 to 25 minutes.

<div align="center">Results</div>

Summary statistics for the four grades tested (grades 2 through
5) on the vocabulary formats and on the reading subtest of the
Metropolitan Achievement Tests are presented in Tables 9 and 10.
Table 11 lists the number of subjects tested on each level of the
three formats. Table 12 summarizes the observed correlational
relationship of vocabulary format to comprehension ability by grade
level. Table 13 presents the results of a two-way analysis of
variance. Examination of the relative size of the correlations
in successive grades on Table 12, reveals that fourth and fifth grade
subjects showed far higher correlations between vocabulary and

Table 9

Summary Statistics for Vocabulary

Formats (% Correct) by Test Level

Test Level	N	$\overline{X}$	SD
1	40	51	26
2	76	51	27
3	109	41	23
4	104	56	28
5	68	45	21
6	30	53	26

Table 10

Summary Statistics for Metropolitan

Comprehension Subtest (Raw Score)

by Grade

Grade Level	N	Md	$\overline{X}$	SD
2	36	32	29.22	10.47
3	33	31	29.38	9.09
4	32	19	19.83	9.13
5	27	23	21.38	8.21

Table 11

Number of Subjects Tested

on Vocabulary Formats by

Level and by Grade

Test Level	Grade				Total
	2	3	4	5	
1	36	0	0	0	36
2	36	33	0	0	69
3	36	33	32		101
4	0	33	32	27	92
5	0	0	32	27	59
6	0	0	0	27	27

Table 12

Pearson Correlations of Vocabulary

Formats with Metropolitan, by Grade

	Formats		
Grade Level	Synonym out-of-context	Synonym in Context	Cloze
2	.408	.422	.435
3	.498	.351	.410
4	.7?7	.889	.717
5	.824	.830	.863

comprehension scores than did second and third grade subjects.
The reasons for this observed split between grades 2-3 and 4-5
are unclear. It is possible that the target words for fourth
and fifth grade test levels were relatively easier for older
subjects than the target words for second and third grade test
levels were for younger subjects. An item analysis of the tests,
however, did not support this conjecture.

A major goal of this study was to determine which vocabulary
format correlated most highly with comprehension within, as well
as across, grade levels. Again, no obvious pattern was apparent.
Unlike the observations in Study 2, primary (second and third)
grade subjects' comprehension scores did not show high correlations
with any of the vocabulary formats, and the differences in per-
formance between the three formats did not appear significant. At
the fourth grade level, the Synonym in Context Test was most
highly correlated with comprehension, but at the fifth grade level,
the Cloze Test had the highest correlation with comprehension. Re-
sults of an analysis of variance showed no significant format
effect and no significant interaction effect of format and grade
(see Table 13). There was, however, a significant grade level
effect ($p < .001$).

Discussion

Study 3 identified two major issues that warrant further in-
vestigation:

1. The vocabulary tests may be of uneven difficulty for

Table 13

Two-way ANOVA of

Main Effects: Format and Grade

Source	df	MS	F	p
Main Effects:				
Format	2	338.46	0.77	0.466
Grade	3	2540.76	5.77	0.001
Interaction:				
Format x Grade	6	211.92	0.48	0.821
Error	116	440.70		

successive grade levels. This notion will be examined in further

administrations of a paper-and-pencil version of the test, with

item analyses to assess test validity at all grade levels.

2. Fourth and fifth grade subjects utilize vocabulary and

comprehension skills in a unified fashion. Assessment methods for

vocabulary knowledge frequently measure some strategies common

to both vocabulary and comprehension abilities. At the primary

grades (grades 2 and 3), subjects' strategies for vocabulary growth

and comprehension appear to be less integrated.

Although the data collected in Study 3 did not resolve the

questions of which vocabulary format most effectively assesses word

knowledge, and which format shows the most consistent performance

pattern across grade levels; Study 3 did suggest an interesting issue

for further investigation: Could the underlying relationship between

vocabulary knowledge and comprehension ability be discontinuous across

grade levels?

Because Study 3 demonstrated that no one of the three vocabulary

assessment formats was significantly inferior to the other two, all

three formats will be included in future studies. The issue of a

possible discontinuity in the relationship of vocabulary knowledge.

to comprehension ability across grade levels will also be explored.

Summary

The results of Study 3 indicated lower correlations between

vocabulary knowledge (on all three assessment formats) and compre-

hension ability for grades 2 and 3 than for grades 4 and 5. No

single format proved to be superior in assessing vocabulary knowledge across all grade levels. A hierarchy of difficulty for the three vocabulary formats was not apparent. Further testing will explore the observed discrepancy of vocabulary and comprehension correlations between the performances of prim ry (second and third) and middle (fourth and fifth) grade subjects. Finally, in future testing sessions, item analyses will be performed on target words to ensure that tests are appropriately graduated in difficulty.

Appendix A

Stim   for Study 1 and Study 2

Self-screening Test, Level 4

Oral Recognition Test, Level 4

Synonym-out-of-context Test, Level 4

Synonym in Context Test, Level 4

Cloze Test, Level 4

Self-screening Test

Level 4

briskly

feeble

hesitate

gigantic

concealed

pauper

confident

menacing

obvious

pursued

Oral Recognition Test

Level 4

Stimulus	Responses		
briskly	briefly	blackie	biscuit
feeble	freedom	table	triple
hesitate	state	hasty	estate
gigantic	gyration	gymnastic	gentle
concealed	concubine	concrete	convinced
pauper	powder	paper	paucity
confident	convince	confetti	condemn
menacing	mendacity	mention	marking
obvious	obtain	abstain	abound
pursued	purse	persist	purpose

Synonym-out-of-context Test

Level 4

Stimulus	Responses			
briskly	slowly	sadly	rapidly	awkwardly
feeble	weak	funny	fleet	strong
hesitate	pity	provide	pounce	pause
gigantic	huge	minute	size	narrow
concealed	famous	displayed	hidden	peaceful
pauper	helpless	unhappy	poor	lost
confident	careful	sure	fearful	hesitate
menacing	inspiring	regular	slinking	dangerous
obvious	evident	crazy	unsure	forgive
pursued	punished	chased	helped	caught

Synonym in Context Test

Level 4

Stimulus	Responses

(The target word is
underlined.)

Stimulus	Responses	
The horse trotted <u>briskly</u> along a familiar path.	slowly rapidly	sadly awkwardly
The <u>feeble</u> old dog walked beside the woman.	weak fleet	funny strong
She did not <u>hesitate</u> in defending herself against the attack.	pity pounce	provide pause
The <u>gigantic</u> giraffe licked the child's hand.	huge size	minute narrow
He <u>concealed</u> the gun in his coat pocket.	famous hid	displayed peaceful
The <u>pauper</u> hand no money for food.	helpless poor man	unhappy lost
The candidate was <u>confident</u> that she would win.	careful fearful	sure hesitant
The robber gave the woman a <u>menacing</u> look after she screamed.	inspiring slinking	regular evil
John's mistake on his math test was an <u>obvious</u> error in subtraction.	evident pleasant	crazy forgive
The cat <u>pursued</u> the mouse across the kitchen floor.	punished helped	chased caught

Cloze Test

Level 4

Stimulus	Response

The horse trotted _____ along
a familiar path.

briskly   after
below   twelve

The _____ old dog walked beside
the woman.

feeble   young
fleet   friend

She did not _____ in defending herself
against the attack.

hesitate   provide
encamp   pity

The _____ giraffe licked the child's
hand.

gigantic   minute
size   blue

He _____ the gun in his coat pocket.

concealed   prayed
peaceful   famous

The _____ had no money for food.

pauper   unhappy
lost   helpless

The candidate was _____ that she would
win.

confident   after
hesitant   heavy

The robber gave the woman a _____ look
after she screamed.

menacing   regular
slinking   inspiring

John's mistake on his math test was an
_____ error in subtraction.

obvious   angry
itching   empty

The cat _____ the mouse across the
kitchen floor.

pursued   smiled
gave   slept

Appendix B

Stimuli for Study 3

Synonym-out-of-context Test, Level 4

Synonym in Context Test, Level 4

Cloze Test, Level 4

Synonym-out-of-context Test

Level 4

Stimulus		Responses		
gigantic	huge	baby	friendly	lonely
gain	need	have	put on	lose
nervous	hungry	uneasy	young	tired
broken	destroyed	shaken	dirtied	moved
concealed	wanted	discovered	found	hid
feeble	funny	unhappy	weak	strong
briskly	slowly	quickly	sadly	awkward
exhibits	acts	people	rides	displays
surprised	amazed	sick	gone	bored
twisted	held	wound	found	placed

Synonym in Context Test

Level 4

Stimulus (The target word is underlined.)	Responses	
The gigantic giraffe licked the child's hand.	huge friendly	baby lonely
Ann will gain weight from eating too much cake.	need put on	have lose
Patty bit her nails because she was nervous.	hungry young	uneasy tired
Many of our dishes were broken by the earthquake.	destroyed dirtied	shaken moved
He concealed the gun in his coat pocket.	wanted found	discovered hid
The feeble old dog walked beside the woman.	funny weak	unhappy strong
The horse trotted briskly along a familiar path.	slowly sadly	quickly awkward
There were many interesting exhibits at the science fair.	acts rides	people displays
Mark was surprised when his friend from Texas arrived.	amazed gone	sick bored
Cathy twisted a rubber band around her finger.	held found	wound placed

Cloze Test

Level 4

Stimulus	Responses

The giraffe was so _____ that he made even the elephants look small.

gigantic young

friendly lonely

Ann will _____ weight from eating too much.

need have

gain lose

Patty bit her nails because she was _____.

smart nervous

young asleep

Many of our dishes were _____ by the earthquake and we had to sweep up all the pieces.

broken shaken

lost moved

He _____ the gun in his coat pocket so no one would know he had it.

made discovered

found concealed

The _____ old dog could barely keep up with the woman, but wagged his tail happily anyway.

quick unhappy

feeble strong

The horse trotted _____ along the path and arrived home sooner than usual.

slowly briskly

sadly awkwardly

There were many interesting _____ at the fair which showed how different foods got from the farm to our tables.

acts people

rides exhibits

Mark was _____ and happy when his friend arrived unexpectedly from Texas.

sick gone

bored surprised

Cathy _____ around so she could reach the book on the shelf behind her.

held twisted

found placed

References

Dale, E. and O'Rourke, J. The living word vocabulary. Elgin, IL, Dome, Inc., 1976.

Venezky, R., Bernard, L., Chicone, S., & Leslie, R. On-line diagnosis of reading difficulties (Technical Report No. 327). Madison: Wisconsin Research and Development Center for Cognitive Learning, 1975.

Venezky, R., Chicone, S., & Perry, J. On-line diagnosis of reading difficulties-II (Technical Report No. 386). Madison: Wisconsin Research and Development Center for Cognitive Learning, 1976.

Venezky, R.L., Perry, J., Chicone, S. and Pittelman, S. Summary of studies for an on-line reading diagnosis system-III (Technical Report No. 409). Madison: Wisconsin Research and Development Center for Cognitive Learning, 1977.

60