DOCUMENT RESUME

ED 090 314 TM 003 603

AUTHOR Hansen, Lee H.; Hesse, Karl D.
TITLE A Criterion-Referenced Assessment of Reading Literacy Using the Cloze Procedure.
PUB DATE Apr 74
NOTE 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, Illinois, April, 1974)

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS *Cloze Procedure; *Criterion Referenced Tests; Elementary School Students; Evaluation Methods; *Literacy; *Measurement Techniques; *Reading Ability; Secondary School Students
IDENTIFIERS Bormuth Literacy Model; Domain Referenced Testing

ABSTRACT

Students at four grade levels (4,7,10, and 12) were tested using a criterion-referenced measure of reading literacy based on the Bormuth literacy model. Cloze test passages were selected to represent a cross section of ten content domains of readers material (e.g., textbooks, consumer safety, occupational, etc.). A multimatrix sampling design was used. Students scoring below the 35 percent correct criterion were judged illiterate with that material. In general, the study suggested far less reading literacy among students in one school system than was anticipated. This paper represents a brief abstract of the complexities of the ultimate design. (Author/BB)

A CRITERION-REFERENCED ASSESSMENT OF READING LITERACY USING THE CLOZE PROCEDURE

Lee H. Hansen

Karl D. Hesse

MADISON PUBLIC SCHOOLS
Madison, Wisconsin
April, 1974

INTRODUCTION

Assessing growth in students' ability to read is a problem that has faced school systems for decades. Most schools have responded by regularly administering standardized tests of reading achievement. Educators and parents alike recognize the inadequacy of these measures, for they do not address a most crucial question, i.e., does my child read well enough to thrive in school and society.

Recently a new approach to this measurement problem was developed by John Bormuth of the University of Chicago (Bormuth, 1971). Bormuth, assuming that reading can at least be defined as gaining information from specific printed messages, developed a test of information gain that utilized a new kind of achievement test item. This measure is a completion type test item that can be systematically generated from a written message using a set of linguistic rules. (Bormuth, 1970b) Such items avoid most of the criticisms normally leveled at other kinds of test items when used in a criterion-referenced test.

As this test of "information gain" was being developed, it became clear that the large developmental cost associated with the measure prohibited its general use as a program assessment tool by school systems. Therefore, Bormuth identified a much cheaper surrogate measure in the cloze test which had many of the same psychometric qualities of the test items for his test of information gain. He found, moreover, that a cloze test and an "information gain" test on the same material had a systematic and highly predictable relationship. More specifically, Bormuth found that it was not until students' cloze

scores reached about 35% that they were able to gain much if any information from what they had read. Thus, a criterion or threshold of minimal reading literacy was suggested for specific reading materials.

This realization presented school systems with an inexpensive assessment model for determining if students can read (can gain information from) what they are expected by society to read. What follows is one school system's attempt to pilot that assessment model under operational conditions. This paper represents but a brief abstract of the complexities of the ultimate design and the reader is referred to the final report for a more thorough description (Hansen and Hesse, 1974).

DESIGN

Question

The question addressed was, "Can students in the Madison Public Schools achieve a correct replacement score greater than 35% on cloze tests developed from samples of reading materials which they (the students) are expected by the community and school to be able to read?"

Instrumentation

Since the cloze tests to be used in the assessment were to be based on reading material which students in our school system were actually expected to read, it was necessary to develop the testing material for the assessment. Briefly, the following procedure was used. With the help of reading specialists and language arts teachers, a taxonomy of reading materials was developed. (Tables 1.1, 1.2, and 1.3) That taxonomy included 10 reading domains: (a) Recreational Literature, (b) School, (c) Automobile, (d) Citizenship, (e) Leisure-time Activities,

(f) Occupational Material, (g) Reference Materials, (h) Safety, (i) Consumer Material, and (j) Textbooks. Spread across these ten domains were 60 categories of material. (See Tables 1.1, 1.2, and 1.3 for the breakdown.)

TABLE 1.1

DOMAINS AND CATEGORIES

		Grade Level Used
A.	Recreational Literature:	
	1. Magazines (3)	4-7-10-12
	2. Stories dealing with family relationships (3)	4-7-10-12
	3. Animal stories (3)	4-7-10-12
	4. Patriotic stories (3)	4-7-10-12
	5. Biographies (3)	4-7-10-12
	6. Adventure stories (3)	4-7-10-12
	7. Student newspapers (3)	4-7-10-12
	8. Mystery stories (3)	4-7-10-12
B.	School:	
	1. Student handbook (3)	7-10-12
	2. Paperback list (15)	4-7-10-12
	3. Standardized test instructions (5)	4-7-10-12*
C.	Automobile:	
	1. Promotional literature on new cars (3)	10-12*
	2. Auto insurance promotion (3)	10-12*
	3. Automotive license manual (3)	10-12*
	4. Automotive driving tips (3)	10-12*
	5. Operator's and owner's manual (3)	10-12*
	6. Penalty point literature (3)	10-12*
D.	Citizenship:	
	1. Wisconsin Constitution (3)	10-12*
	2. Voting directions (3)	10-12*
	3. Newspapers (40)	7-10-12*
	4. Referenda statements (27)	10-12*

( ) Number indicates the quantity of passages which will represent that category.

* Indicates all grades involved received the same passages.

TABLE 1.2

DOMAINS AND CATEGORIES

		Grade Level Used
E.	Leisure-time Activities:	
	1. Rule books (sports) (3)	7-10-12*
	2. Directions for assembling toys (3)	4-7-10-12*
	3. Recreation department bulletins (3)	4-7-10-12
	4. Directions for games (3)	4-7-10-12*
	5. T.V. Guide (3)	4-7-10-12*
	6. Boy/Girl Scout Manual (3)	4-7-10-12*
	7. Directions accompanying sewing patterns (3)	4-7-10-12*
F.	Occupational:	
	1. Vocational School annual ad (3)	10-12*
	2. Instructions on job applications (3)	10-12*
	3. Civil Service test instructions (3)	10-12*
	4. Armed Forces promotional literature (3)	10-12*
	5. Prospective careers promotional literature (3)	10-12*
	6. Instructions for filling out job-related forms (3)	10-12*
	7. School catalogs (3)	10-12*
G.	Reference Materials:	
	1. Road maps (3)	7-10-12*
	2. Telephone directory (3)	4-7-10-12*
	3. Encyclopedia (3)	4-7-10-12
	4. Reference books (3)	4-7-10-12
H.	Safety:	
	1. Fire Department literature (3)	4-7-10-12*
	2. Airplane emergency literature (3)	4-7-10-12*
	3. Civil Defense instructions (3)	4-7-10-12*
	4. Bicycle and pedestrian rules and fire evacuation (5)	4-7-10-12*
	5. Warnings and instructions on commercial packaging (10)	4-7-10-12*
	6. Heart, Cancer, and Red Cross literature (3)	7-10-12*
	7. Directions for power tools (3)	7-10-12*
I.	Consumer Material:	
	1. Junk mail (3)	7-10-12*
	2. Messages on packages (3)	4-7-10-12*
	3. Recipes (3)	7-10-12*
	4. Catalogs (3)	7-10-12*
	5. Contractual agreements (3)	10-12*
	6. Banking promotional literature (3)	10-12*
	7. Financial planning literature (3)	10-12*
	8. Advertisements (3)	7-10-12*
	9. Conservation/ecology literature (3)	7-10-12*
	10. Consumer magazines (3)	10-12*

TABLE 1.3

DOMAINS AND CATEGORIES

J. Textbooks:	Grade Level Used
1. Language Arts and Reading (12)	4-7-10-12
2. Social Studies (12)	4-7-10-12
3. Science (12)	4-7-10-12
4. Mathematics (12)	4-7-10-12

Within each category of each reading domain a set of documented decision rules was used to obtain representative messages for each category. After these messages had been gathered and appropriately coded, a random 70 word passage was extracted from each message. Since analysis of the data was not to be extended to the message level, it was not necessary to sample multiple passages for each message. All of the passages for a particular reading domain made up a reading subtest for that domain. Subtests across each domain made up the total test. Each 70 word passage was clozed by eliminating every fifth word and substituting a uniformly sized line until a total of 10 cloze test items was identified for each passage.

Population and Sample

Students in grades 4, 7, 10, and 12 were selected for the assessment. A random 50% sample of the fourth graders was tested; all seventh, tenth, and twelfth graders were tested.

Design

It is evident from examining the categories and domains in Tables 1.1, 1.2, and 1.3 that society may not have the same expectations of students at all four grade levels. Thus each category within each domain was

carefully screened to determine the lowest of the four grade levels at which students were expected to read the material. This was accomplished by submitting the categories within each domain to a panel of judges made up of parents, teachers, and administrators. Moreover, it had become quite apparent when messages were being collected that for some categories and domains, e.g., textbooks or leisure reading, the difficulty of reading material was controlled for each of the four grade levels. This was accounted for when messages were sampled. Thus, not all domains and categories were tested with all four grade levels, and some categories and domains contained different reading material at each grade level.

Tests for each grade level were constructed by identifying the passages within each domain and category that were to be administered. These passages were in turn randomly assigned to a series of test forms (the exact number of forms varied with the grade level). A multi-matrix sampling design was then employed to randomly assign one of these test forms to each student; thus each student did not respond to all passages in the total test, but rather to a random subsample of all possible passages. Table 2.0 displays the number of passages, test forms, and students (on the average) responding to each test form.

TABLE 2.0

MULTI-MATRIX SAMPLING PLAN FOR EACH GRADE LEVEL

Grade Level	4	7	10	12
Total Passages	144	216	300	300
Population	2500	2400	2300	2100
Students Per Passage	60	60	60	60
Passages Per Booklet Form (Including Common Passage)	9	7	11	11
Number of Booklet Forms	18	36	30	30
Overall Sample Needed	1080	2160	1800	1800

A common passage was also included in each test booklet. This passage was used to test the validity of the multi-matrix sampling design.

Data Collection and Preparation

Tests were administered by a cadre of trained test administrators. All testing occurred in a single, untimed setting. Students were allowed to get help from the test administrator in spelling words if they could verbally supply the word they wished spelled.

Tests were scored by determining the percentages of cloze items for which the eliminated word had been successfully replaced by each student. Only exact words rather than synonyms or grammatically similar words were scored correct because that was consistent with Bormuth's assessment model.

Results

The original analysis design called for aggregating by domain the results obtained for passages tested on different subsamples of students at

each grade level. However, this is a defensible procedure only if the subsamples of students drawn randomly from the total population are statistically equivalent. This hypothesis was tested using as a dependent measure a ten item cloze passage commonly administered to all students. A One-Way Analysis of Variance was conducted among the subsamples at each grade level. The F-values for these tests are presented in Table 3.0.

TABLE 3.0

F-RATIOS FOR ONE-WAY ANOVA TESTS CONDUCTED ON A COMMON PASSAGE SCORE AMONG SUBSAMPLES OF STUDENTS AT EACH GRADE LEVEL

Grade	DF	F-Ratio
Grade 4	17, ∞	1.11
Grade 7	35, ∞	1.25
Grade 10	29, ∞	.73
Grade 12	29, ∞	1.29

None of these values is significant at the .10 level; therefore subsamples of students at each grade have a high probability of being statistically significant. Thus, it was possible to aggregate by domain results obtained on subsamples of students. These data are presented in the paragraphs that follow.

Tables 4.1 through 4.4 display data for the four grade levels. Percentages of students who scored above and below the minimal literacy criterion of 35% are portrayed for each reading domain tested at that given grade level.

PERCENTAGES OF STUDENTS SCORING ABOVE AND BELOW
THE 35% CRITERION LEVEL FOR EACH READING DOMAIN SUBTEST

TABLE 4.1 - GRADE 4

Domain	Below 35%	35% and Above
Recreational Reading	66%	34%
School-Related	48	52
Leisure-time Activities	82	18
Reference	62	38
Safety	84	16
Consumer	77	23
Textbook	65	35

TABLE 4.2 - GRADE 7

Domain	Below 35%	35% and Above
Recreational Reading	62%	38%
School-Related	44	56
Citizenship	73	27
Leisure-time Activities	54	46
Reference	64	36
Safety	48	52
Consumer	60	40
Textbooks	47	53

TABLE 4.3 - GRADE 10

Domain	Below 35%	35% and Above
Recreational Reading	34%	66%
School-Related	37	63
Automobile	50	50
Citizenship	54	46
Leisure -time Activities	42	58
Occupational	45	55
Reference	57	43
Safety	23	77
Consumer	42	58
Textbooks	31	69

TABLE 4.4 - GRADE 12

Domain	Below 35%	35% and Above
Recreational Reading	28%	72%
School-Related	32	68
Automobile	48	52
Citizenship	44	56
Leisure-time Activities	34	66
Occupational	33	67
Reference	46	54
Safety	17	83
Consumer	32	68
Textbooks	26	74

Data in Table 4.1 suggest that large numbers of fourth grade students were unable to correctly fill in as many as 35% of the cloze blanks irrespective of the domain of reading material on which they were tested. Even in the domain labeled School-Related, where fourth graders had the highest scores, 48% of the students had cloze scores below 35%. Results for domains such as Leisure-Time Activities, Safety, and Consumer, where percentages of students below the 35% criterion were 82%, 84%, and 77% respectively, suggest that in general fourth graders are unable to gain information from these materials. In addition to School-Related, fourth graders scored the highest on materials such as Recreational Reading, Reference, and Textbooks where some attempt has been made by teachers to match the level of difficulty of the reading material with the general ability of fourth graders.

Data found in Table 4.2 suggest that at the seventh grade level there are, as at the fourth grade level, substantial numbers of students, irrespective of the reading material domain, who are unable to fill in as many as 35% of the blanks in the test passages. The percentages run from a low of 44% for School-Related materials to a high of 73% for Citizenship materials (primarily newspapers at the seventh grade level). Notably, too, the number of seventh graders who could not complete 35% or more of the blanks of test passages from their textbooks (47%) is somewhat lower than was true at the fourth grade level. Whereas two out of every three fourth graders could score no higher than 34% correct in test passages from textbooks assigned to them, one out of every two seventh graders could score above 35% on test passages from textbooks assigned to them.

Again, it should be noted that two of three domains where seventh grade students performed the best are related to the process of schooling (School-Related and Textbooks) where some efforts are made to select materials which the average seventh grader can supposedly read. However, in spite of this screening, better than 40% of the seventh graders are still unable to gain information from these materials.

The data in Tables 4.3 and 4.4 suggest that at grades 10 and 12 the scores divide more clearly between readers and non-readers. Irrespective of the domain examined, at least 56% of the high school population are scoring 35% or more correct, thus suggesting that they can gain at least some information from the wide range of material represented in the study. However, disturbingly large percentages of 10th and 12th graders are still scoring below 35% correct and thus must be considered illiterate with the material in question.

An analysis of differences among the various domains of material suggest that relative material difficulty is a larger variable in accounting for differences in level of performance at grades 10 and 12. At both grades, for example, students appear to have somewhat better scores on Recreational Reading, School-Related, Safety, and Textbooks than they have for categories relating to Automobile, Citizenship, Leisure-Time Activities, Occupational, Reference, and Consumer. Materials in the domains of Citizenship and Reference materials appeared to be particularly difficult to 10th graders, since 54% and 57% of the students in those respective domains had scores below 35%. Twelfth graders also had considerable difficulty with those two domains with 44% and 46% scoring below the 35% criterion of literacy.

Finally, since both 10th and 12th graders responded to tests of the same passages of material, it is possible to draw direct comparisons between the performance of the two groups on each of the domains. Except for the domain entitled Automobile, anywhere from 5 to 12% fewer 12th graders had scores below 35% than did 10th graders. For Automobiles performance is virtually unchanged (50% vs. 48%). The largest reductions in percentages of students below 35% occurred in those domains that proved particularly difficult for 10th graders: Citizenship, Occupational, Leisure-Time Activities, Reference, and Consumer.

Discussion

The Dormuth model for assessing reading literacy postulates explicit standards for judging whether or not students can read well enough to meet the expectations of society, including the schools. If these standards are accepted as reasonable, several general observations can be made and discussed from the data reported above.

First, it appears evident that substantial numbers of students (somewhere between 25 and 30%) leave high school unable to gain much if any information from the basic material society is expecting them to read. Only in the case of safety materials was the percentage of 12th grade students with scores below 35% as low as 17%.

One's immediate reaction to such a conclusion is disbelief. Surely the repercussions of having a society where virtually 25% or more of the people have difficulty gaining information by reading would be greater than it is. That may well be true. On the other hand, the fact that the repercussions are no greater than they are can perhaps be attributed to another hypotheses: most poorly reading adults are able to make

reasonable compensations for their handicap in this media rich society. Reading may not, in fact, be as economically important a skill as some have claimed it is. Perhaps what is needed is an inventory of adult reading habits and needs to determine how literate people have to be in order to prosper in society.

A second observation that can readily be made from the data is that children's reading power develops much more slowly than most educators and lay people would predict. Virtually 68% of the 4th graders and 56% of the 7th graders had scores below the 35% criterion level irrespective of the domain. Even at grade 10 as many as 30 to 50% of the students had scores below the 35% criterion threshold of literacy, although at both grades 10 and 12 these percentages were much more dependent upon the domain of material in question.

Even on categories and domains where the material, although similar in content, is written to the expectations of the grade level, the growth in reading power is not noticeable. For example, on Recreational Reading 66% of the 4th graders, 62% of the 7th graders, 34% of the 10th graders, and 28% of the 12th Graders were below 35%. The perceptions of those authors who write to audiences at each of the four grade levels would appear to be unrealistic; students would not appear to have the level of reading power expected by these authors.

A third inescapable observation is that considerable numbers of 4th, 7th, 10th, and 12th grade students would appear to be unable to gain much information from the textbooks and reference materials purchased for their use in school, although in the case of textbooks and other school-related materials there is evidence of somewhat better control of readability than is true of non-school related materials.

If this is indeed true, it represents a major problem for educators. Teachers have long sensed the problem of having to work with materials that far too many of their students find exceedingly difficult to read. Moreover, they have tended to seek ways to compensate for the problem as, for example, by lecturing on the material to be read, by introducing vocabulary, by having the material read aloud in class, by giving questions in advance, or by using additional aids such as films and filmstrips to supplement the reading assignment.

The problem also presents dilemmas for educators who have the responsibility of purchasing instructional materials and for textbook companies who have a pecuniary interest in selling their materials. Certainly teachers and administrators must be troubled by expenditures for printed instructional materials when substantial numbers of students can apparently gain little or no information from those materials. On the other hand, there may be no immediate way out of the dilemma, since more readable materials which maintain grade level interest patterns are generally not available from textbook companies who at best have taken a rather lackadaisical stance toward textbook materials readability.

Summary

The data in this report suggest that substantial numbers of students may be leaving schools unable to gain information from what society may expect them to read. Such data are not inconsistent with what others have found when using the same technique. Bormuth (1970a), for example, found evidence "showing that 30 to 60 percent of the students in our better suburban schools read so poorly that they are able to learn little or none of the information contained in their textbooks." Gundlach (1973) obtained

similar results in a literacy assessment conducted in several small communities in Wisconsin. Informally, many teachers, but particularly high school teachers, have with great candor assured these writers that the results are not unrealistic, but do in fact reflect the reading problems which far too many high school students have.

Finally, this situation is not unique to Madison. Results from standardized reading tests suggest that Madison Public School students generally perform above national norms. If these national norming samples are at all representative of children across the country, then it would seem possible with some caution to suggest that other communities will get similar literacy results. Perhaps these results dramatize the severity of a national reading problem that many citizens are only beginning to become aware of.

BIBLIOGRAPHY

Bormuth, John R. Illiteracy in the Suburbs. October, 1970a. (Unpublished paper, University of Chicago).

Bormuth, John R. On the Theory of Achievement Test Items. Chicago, Illinois: The University of Chicago Press, 1970b.

Bormuth, John R. Development of Standards of Readability: Toward a Rational Criterion of Passage Performance. Final Report, Project No. 9-0237. Bureau of Research, U.S. Office of Education, 1971.

Gundlach, Paul L. "Administrative Accountability as Determined by Percentage of Functional Literate Students." Ph.D. Dissertation, University of Wisconsin, October, 1973.

Hansen, Lee H. and Karl Hesse. A Pilot Reading Literacy Assessment of Madison Public School Students. Madison, Wisconsin Public Schools, March, 1974.