DOCUMENT RESUME

ED 071 046 24 CS 000 305

AUTHOR	Tuinman, J. Jaap
TITLE	Obtaining Indices of Passage Dependency of Comprehension Questions. Final Report.
INSTITUTION	Indiana Univ. Foundation, Bloomington.
SPONS AGENCY	National Center for Educational Research and Development (DHEW/OE), Washington, D.C.
BUREAU NO	BR-2-E-005
PUB DATE	15 Oct 72
GRANT	OEG-5-72-0026(509)
NOTE	89p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Elementary Grades; Grade 4; Grade 5; Grade 6; Reading; *Reading Comprehension; Reading Materials; *Reading Research; Reading Skills; *Reading Tests; Testing Problems; *Test Interpretation; Test Results

ABSTRACT

Tests of reading comprehension presently used do not provide one important item of technical data: the extent to which questions used in the test could be answered without reading the paragraphs upon which those questions are based (paragraph dependency). This leaves the test user guessing as to whether the students taking the test and performing well did or did not understand the written material contained in the test. Indices of paragraph dependency for five widely used standardized tests of reading comprehension were obtained. Five tests were administered to 1200 students each, not allowing these students to read passages. In addition, control data were obtained by administering the tests in their normal format to 600 students each. Students were selected from 10 locations covering Indiana and were equally divided over grades 4, 5, and 6. The results indicated that none of these major tests provides sufficient guarantees against the answering of items on the basis of information other than that presented in the passage. Average probabilities of correct responses with no passage present ranged between .32 and .50, well above the expected chance score of .25. (Author)

Final Report

Project No. 2-E-005
Grant No. OEG-5-72-0026 (509)

Obtaining Indices of Passage Dependency of Comprehension Questions

J. Jaap Tuinman

Indiana University

Bloomington, Indiana

October 15, 1972

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
National Center for Educational Research and Development

Preface

Though I am convinced that, in the long run, our schools stand to benefit from the work done by so many educational researchers, it is the latter who, in most cases, benefit most directly from any cooperation between the two. I am therefore deeply indebted to the personnel in the ten Indiana school systems that participated in this study and to the students in their schools. I wish to thank all of them. In particular, I am grateful to my primary contacts in each of the school systems: Mrs. Helen McDaniel and Messrs. Leo Joint, Donald Eberly, David Whaley, Donald Massey, Herbert Reese, Melvin Mozier, George Westfall, Charles Arvin and J. O. Smith.

Collecting as much data as we did is impossible without the underpaid help of many graduate students. I hope this enterprise was an education for at least some of them. Travelling across Indiana with big boxes of testing materials was most of the time hard work and only sometimes fun. I wish to thank my regular crew: Carol Brooks, Dave Downing, Beverly Farr, P. J. Fitzgerald, Mary Halpin, Linda Hoyman and Dick Szuny.

Finally, what would I have done without the diligent work of Mary Ella Brady, Jeanne Burns and Mary Halpin in the analysis, write up and run-off stage? Nothing!

J.J.T.

Table of Contents

Preface i

Table of Contents ii

List of Tables iii

List of Figures vii

Introduction 1

Subjects 12

Procedure 15

Results 18

Discussion 35

References 37

Appendix A 39

Appendix B 46

Appendix C 53

List of Tables

Table		Page
1	Number of Ss Per School System Administered Tests Under Passage (P) and No Passage (NP) Conditions	16
2	Means, S.D.'s, KR-20 Coefficients and S.E.'s of Measurement for P and NP-Conditions Across Grades	19
3	Means Under the NP-Condition Expressed as (1) %ages of the Number of Test Items and (2) As %ages of the Means Obtained Under the P-Condition	20
4	Test 1 - Means, S.D.'s, KR-20 Reliability Coefficients and S.E.'s of Measurement for P and NP-Conditions. Results by Grade	22
5	Test 1 - Means Under the NP-Condition Expressed (1) as %ages of the Number of Test Items and (2) as %ages of the Mean Under the P-Condition. Results by Grade	22
6	Test 2 - Means, S.D.'s, KR-20 Reliability Coefficients and S.E.'s of Measurement for P and NP-Conditions. Results by Grade	23
7	Test 2 - Means Under the NP-Condition Expressed (1) as %ages of the Number of Test Items and (2) as %ages of the Mean Under the P-Condition. Results by Grade	23
8	Test 3 - Means, S.D.'s, KR-20 Reliability Coefficients and S.E.'s of Measurement for P and NP-Conditions. Results by Grade	24
9	Test 3 - Means Under the NP-Condition Expressed (1) as %ages of the Number of Test Items and (2) as %ages of the Mean Under the P-Condition. Results by Grade	24
10	Test 5 - Means, S.D.'s, KR-20 Reliability Coefficients and S.E.'s of Measurement for P and NP-Conditions. Results by Grade	25
11	Tests 4 and 5 - Means Under the NP-Condition Expressed (1) as %ages of the Number of Test Items and (2) as %ages of the Mean Under the P-Condition. Results by Grade	25

Table		Page
12	Test 6 - Means, S.D.'s, KR-20 Reliability Coefficients and S.E.'s of Measurement for P and NP-Conditions. Results by Grade.	26
13	Test 6 - Means Under the NP-Condition Expressed (1) as %ages of the Number of Test Items and (2) as %ages of the Mean Under the P-Condition Results by Grade	26
14	Number of Items Per Test with a Difficulty Under the NP-Condition Higher than 1/k, Where k = the Number of Options Per Item.	29
15	Percentage of Items Per Test with a Difficulty Under the NP-Condition Higher than 1/k, Where k = the Number of Options Per Item	30
16	PDI_2 Values for Tests 1-6; By Grade	33
17	Average Difficulties Under P and NP Conditions and E_2 Values for Six Tests; Combined Across Grades	34
A1	Test 1 - Item Difficulties Under the NP-Conditions, Per Grade	40
A2	Test 2 - Item Difficulties Under the NP-Conditions, Per Grade	41
A3	Test 3 - Item Difficulties Under the NP-Conditions, Per Grade	42
A4	Test 4 - Item Difficulties Under the NP-Conditions, Per Grade	43
A5	Test 5 - Item Difficulties Under the NP-Conditions, Per Grade	44
A6	Test 6 - Item Difficulties Under the NP-Conditions, Per Grade	45
B1	Test 1 - Difficulty Coefficients Under P and NP-Conditions (d_P, d_{NP}), Passage Dependency Index 2 (PDI_2) and Passage Dependency Efficiency Index (E_2) Combined Across Grades	47

Table		Page
B2	Test 2 - Difficulty Coefficients Under P and NP Conditions (d_P, d_{NP}), Passage Dependency Index 2 (PDI_2) and Passage Dependency Efficiency Index (E_2)	48
B3	Test 3 - Difficulty Coefficients Under P and NP Conditions (d_P, d_{NP}), Passage Dependency Index 2 (PDI_2) and Passage Dependency Efficiency Index (E_2)	49
B4	Test 4 - Difficulty Coefficients Under P and NP Conditions (d_P, d_{NP}), Passage Dependency Index 2 (PDI_2) and Passage Dependency Efficiency Index (E_2)	50
B5	Test 5 - Difficulty Coefficients Under P and NP Conditions (d_P, d_{NP}), Passage Dependency Index 2 (PDI_2) and Passage Dependency Efficiency Index (E_2)	51
B6	Test 6 - Difficulty Coefficients Under P and NP Conditions (d_P, d_{NP}), Passage Dependency Index 2 (PDI_2) and Passage Dependency Efficiency Index (E_2)	52
C1	Test 1 - Values of Item Validity Statistics Described in Appendix C for all Items	71
C2	Test 2 - Values of Item Validity Statistics Described in Appendix C for all Items	72
C3	Test 3 - Values of Item Validity Statistics Described in Appendix C for all Items	73
C4	Test 4 - Values of Item Validity Statistics Described in Appendix C for all Items	74
C5	Test 5 - Values of Item Validity Statistics Described in Appendix C for all Items	75
C6	Test 6 - Values of Item Validity Statistics Described in Appendix C for all Items	76
C7	Illustrative Validity Statistics for Selected Values of p_a and p_c	77

Table		Page
C8	Mean Values of Validity Statistics for Six Tests	78
C9	Validity Statistics for Selected Items	79
C10	Correlation Matrix - Test 5 (N = 45)	80

List of Figures

Figure		Page
1	State of Indiana	13
2	School Systems, Number of Schools, and Students Per System	14

INTRODUCTION

The purpose of this study was to determine to what extent items in a number of selected standardized tests of reading can be answered without prior reading taking place. Indices of paragraph dependency were calculated. The study was limited to so called tests of paragraph and/or story comprehension.

Tests of reading comprehension purport to measure how well a student understands what he is reading. Many of these tests employ questions to ascertain the degree of this understanding. This technique is based on the tacit assumption that a direct relationship exists between the reading of the passage or story and the answering of questions about it. In the case of a great many reading test items from standardized tests this is a faulty assumption.

More than 25 years ago Davis (1944) asked the question: "What do reading tests really measure?" His answer to his own question - an answer repeated in essence in his more recent study (Davis, 1968) - indicated a definite dissatisfaction with the inability of standardized tests of reading to measure the skills "considered highly important by the authorities in the field (Davis, 1944, p. 187)."

The legitimacy of a concern for the functioning of reading tests as they are now known is underscored by a series of recent studies which revealed that in many cases successful performance on the reading measure was only loosely related to the necessity for the reader to have read the passage on which the questions presumably were based.

A relatively detailed picture of students' ability to answer comprehension questions without the aid of the text from which they are derived is provided by studies by Weaver and Bickley (1967), Bickley, Weaver and Ford (1968) and Weaver, Bickley and Ford (1969). In a series of studies utilizing the black-out technique, one of the recurrent experimental conditions was that students were required to answer multiple-choice items sampled from reading tests listed in the Sixth Mental Measurement Yearbook (Buros, 1965), with the accompanying reading passages completely blacked out. In light of what one conventionally assumes about the function of a reading test, their finding is somewhat startling: "The Ss who had no reading passage to aid in answering the items, nevertheless, correctly completed 67% as many items as Ss with all the reading passage" (Weaver and Bickley, 1967, p. 294). A further analysis of this phenomenon led these authors to conclude:

> In other words, with the materials here, there is a difference between having or not having a reading paragraph, even in the less relatedness of reading paragraph condition, but this effect is much more pronounced in the more relatedness to reading paragraph condition. (Weaver, Bickley and Ford, 1969, p. 12)

The above statement may be interpreted to mean that items of a relatively more factual nature, to be answered directly on the basis of information in the passage, are easier to answer without the paragraph present than are items which are only indirectly related to the information in the paragraph - the inferential items.

Weaver and Bickley (1967) suggest a number of possible ways in which the Ss could have answered the test items without aid of the relevant passage: knowing the answer from prior learning; elimination of irrelevant distractors; the use of information embedded in preceding items. Samuels (1968) demonstrated that high associations among elements in an item stem and the correct distractor, too, facilitate answering of reading comprehension items prior to reading the passage.

In order to evaluate the consequences of the above findings, it is necessary to recall the distinction between a necessary condition and a sufficient condition (Carney and Scheer, 1964, p. 207). It can easily be granted that reading involves relating whatever is being read to prior experience. As such, prior learning is a necessary condition to reading, as long as a definition of reading includes a reference to understanding. The statement that prior learning plays a legitimate role in the answering of multiple-choice questions subsequent to having read a passage is quite acceptable. However, the fact that prior learning is a necessary condition for answering these items does not make it a sufficient condition. A reading test, for instance, is distinguishable from a listening test: in addition to the prior learning and knowledge present, some reading, and not some listening, takes place before the test is taken. In short, it seems reasonable to require that a reading test measure sets of behaviors which are functionally related to reading a passage.

It must be clear that any measure of some variable operates best when irrelevant sources of information germane to performance on that measure are eliminated. Of the three sources of information listed by Weaver and Bickley (1967), none seems exclusive to reading tests. However, whereas the elimination of the last two sources (irrelevant distractors and related items) may require strategies common to test construction in general, the control of the first source (prior learning) may be achieved in a way relatively unique to the area of reading tests.

Reading, as a skill-centered area of instruction, is relatively content independent. That is, learning to read does not primarily mean to acquire a body of knowledge but rather to master a set of skills. As a consequence, the maker of reading tests is relatively free of the obligation to have his tests represent information which embodies existing knowledge in a given area of human studies. In principle, there is no reason why reading skills cannot be tested with materials which actually represent modifications of commonly accepted statements of relationships between elements of reality. It is this freedom in the construction of reading tests which allows the test constructor to control the influence of past learning to a greater extent than is possible in most other areas of testing for scholastic achievement.

From the studies reviewed above, it becomes clear that, in few cases, test authors have been able to capitalize on this characteristic of reading measures. There is a great deal of evidence that the lack of passage control found by Weaver <u>et al</u>. is not limited to the test items which happened to be selected into their instruments.

Preston (1964) had 128 college freshmen take the first 30 comprehension items of the Cooperative English Test: Test C2, Reading Comprehension (Higher Level), Form R, without the passages which the items were supposed to test. After taking the passageless test, <u>S</u>s took the test in the conventional way. On the 30 items the expected mean score was 6. The obtained score was 8.34 ($p < .001$). A second interesting finding was that the ability to answer questions without passages had a low correlation with scores on the regular administration of the test (r = .20) and none at all with vocabulary (r = .13, n.s.).

Bloomer and Heitzman (1965) report findings which tend to substantiate this low correlation found between answering questions with relevant information present and with that information absent. In their experiment, a group of eighth grade students took a pretest consisting of multiple-choice questions, then read the reading passages and took the pretest as a post-test. The correlation between the scores was .12 which was not significant (n=36). During the post-test, however, the information was present only to the extent the student had memorized it.

The study by Christensen and Stordahl (1955) is an example of the difficulties in researching reading comprehension that arise from the fact that the reading of the passages sometimes adds relatively little information. Their research attempted to determine the relative effectiveness of various organizational aids in comprehension and retention. In all, 36 treatments were administered with 12 subjects per treatment. Subjects were Air Force trainees. No significant differences were found among treatment group post-test means. The experiment, replicated with another reading passage, resulted in another set of nonsignificant differences. In their attempts to find an explanation for the results, the authors touched upon the possibility that something might have been wrong with their materials. They did not, however, compare pre- and post-test means. On the passage for which they reported detailed data the overall pretest mean was 88% of their post-test mean, indicating that hardly any information was gained by reading the passages. The tests, with a mean item difficulty near 50%, were of reasonable difficulty.

Their research seems to have been aborted by the nature of the questions. Those questions did not require reading as a necessary condition. In this study of comprehension, therefore, behaviors were studied which were under relatively little control of the reading passages. The extent of this control, apparently, was a variable of unknown quantity in this study.

While the Weaver, et al. studies mentioned above were done with college students, Mitchell (1967) got comparable results with fourth grade pupils using a different test (Gates Basic Reading Test). Noteworthy in Mitchell's study is that boys with low I.Q.'s scored no worse on the "passage-out" items than they did on a test which included the passage.

About 40 years ago Eurich (1931) grappled with the issue basic to the present study. He constructed two reading passages with 50 multiple-choice items each. Passage A was of a general nature, whereas passage B contained highly specific and exact material. The first observation of interest to the present discussion made by Eurich is that while for "after reading" the reliabilities of the two tests were in the same order of magnitude, they differed vastly for the "before reading" condition, with the coefficient for the B passage being very low. Seemingly, the nature of the content of the passages largely determined the results under the "no-passage" condition. No uniform conclusion regarding the function of test items under that condition seemed possible in Eurich's case. (Here, as before, one must keep in mind that the "after reading" condition does not imply actual presence of the passages while the items were being answered.) Further information of interest is the correlation between "before" and "after" reading performance. For Eurich's test A this correlation equaled .37; for test B, .45. Thus, only between 13 and 20% of the variance of the scores before and after was accounted for by a common factor. Unlike Christensen and Stordahl's (1955) study, Eurich's data revealed large mean differences between pre and post-tests.

Tuinman (1970), as part of a study involving experimental items designed to be highly passage dependent, administered the first 40 items of the Sequential Test of Educational Progress - Reading, Form 3A. The mean score obtained by 134 7th, 8th and 9th graders was 20.06 when the passages were presented and 13.66 when only the questions were given. Thus, the "passage-out" score was 34% of the possible score and 68% of the score under the "passage-in" condition.

Farr and Smith (1970) administered 32 items from the Nelson-Denny comprehension test to college sophomores and students. Initially the items were administered without the paragraphs. After a 3-week interval a retest followed with the paragraphs present. They found that for five of the items the number of correct responses under the "passage-out" condition exceeded the number of right answers under the "passage-in" condition. Also, for 12 of the items the number of correct answers in the "passage-out" condition exceeded 50%.

The studies reviewed above indicate that quite a few items on standardized tests have little passage dependency. The item that has a response probability in the passage-out condition of 1/k, where k = the number of options, is rare indeed. Per force, the same holds true for the test whose mean score equals 1/n, where n = the number of items when only the test items and not the passages are being administered.

Does this mean that therefore such items and such tests are invalid and of little use? Not necessarily. Lack of passage-dependency signals potential invalidity more than actual lack of validity. It must be clear that if indeed an item is responded to without prior reading of the text or paragraph, that item constitutes an invalid measurement in the context of a reading comprehension test. However, from the fact that an item is answerable without such prior reading of the text does not follow automatically that Ss taking the test will indeed not read the text. For this reason low passage dependency is "merely" a threat to valid measurement and not proof of invalidity.

In the light of the above comments it becomes of some importance to determine whether children indeed are tempted to skip paragraphs when taking reading tests. Recently, an attempt was made to ascertain to what extent children will engage in such potentially test invalidating behavior as partial or complete passage skipping.

In the first study (Tuinman, 1972a), 60 sixth graders were randomly assigned to one of four treatment groups having to read long passages (L) or short passages (S) paired with either passage dependent (D) or passage independent (I) questions. Thus, four treatment test booklets were constructed (LI, SI, LD and SD). The short passages were incorporated in the long ones. The mean passage dependency of I questions was .58; that of the D items was .25. These statistics were obtained during a pilot study.

The test booklets contained 20 cardboard pages. On the front of each page was a question, on the back of it the accompanying story. Ss were told to take the test in any fashion they wanted. The dependent variable of interest was the number of items answered without a single glance at the passages. Whereas the effect of passage length was not significant, the effect of item type was. The Ss skipped significantly more I-items than D-items.

A second study (Tuinman, 1972b) employed the same stimulus materials in a slightly different experimental design. First, a time pressure variable was added. ("There is a time limit" vs. "There is no time limit"). Secondly, the potential effect of an artificial "set" due to long sequences of highly passage dependent items or highly independent items was reduced by using a repeated measure design. To each subject a set of mixed I and D-items was administered. Again the I-items invited more passage-skipping. Though the mean "skip" score was low (2.5 out of a possible 16) the range of scores (0-10) indicated that individual students may well invalidate their test and (in the case of I-items) get away with it.

From the above discussion it becomes quite clear that (1) individual students may produce responses which are not under control of the passage and (2) that standardized reading tests contain many items which reward rather than punish such behavior.

In the past, test authors and publishers have given little attention to passage-dependency. Its desirability has been only sporadically stressed by test reviewers. The intent of the current study therefore is threefold.

First, attention is called to the degree of lack of passage dependency by obtaining data on five major reading tests.

Secondly, an attempt is made to produce reliable item validity statistics (in particular, passage dependency indices) by using samples larger than those used in most of the research reviewed above.

Thirdly, the shift in passage dependency of items and tests as a function of educational growth of the respondents is demonstrated by selecting Ss in three consecutive grade levels.

PROCEDURE

Tests

Tests were selected for analysis in terms of passage dependency based on the following criteria:

a. Comprehension should be measured by means of the passage-questions technique.

b. Preferably one level of the test would be suitable for administration in grades 4 through 6.

c. The length of the tests would allow students to finish within one hour.

d. The test should be widely used on a national level.

The final selection of tests used in the present study was as follows:

Test 1 - Nelson Reading Test, Form A

Number of items: 75

Test 2 - California Achievement Tests, Level 3 - Form A

Number of items: 42

Test 3 - SRA - Achievement Series, Reading, Form E, Blue level

Number of items: 60

Test 4 - Metropolitan Achievement Tests, Reading - Elementary Battery Form F. Number of items: 45.

Test 5 - Metropolitan Achievement Tests, Reading - Intermediate Battery, Form F. Number of items: 45

Test 6 - Iowa Test of Basic Skills - Reading, Multilevel, Form 5. Number of items: 60

This list of tests requires some comments. First of all, it may be noted that tests 4 and 5 are actually only two different levels of the same test. This is a function of the fact that the Metropolitan did not meet criterion b: no one level of this test was suitable for grades 4, 5 and 6. Therefore, it was decided to use the Elementary Battery with the 4th grade and the Intermediate Battery with the 5th and 6th grades. A second comment which needs to be made regards tests 3 and 6. The multilevel SRA contains far more items suitable for use in grades 4, 5 and 6 than can be answered within one hour. For this reason Test 3 constitutes a subset of SRA items. This subset was arrived at by random selection from the pool of suitable passages of as many passages as were needed to construct a reasonably long test. This procedure resulted in the inclusion of 60 items in Test 3. A similar procedure was followed for Test 6.

Experimental versions* of the tests were created by mimeographing the passages and the items separately. Thus, each test consisted of a passage-booklet and a question-booklet. In the question booklet references to the passage booklet were made that indicated which passage should be read with

---

* The author wishes to thank Harcourt, Brace, Jovanovich, Inc., Science Research Associates, Inc.; CTB/McGraw Hill, Inc. and Houghton Mifflin for permission to use their tests in this research.

which items. The items were left intact, with the exception of changes made necessary by the different print format of the passages. For instance, instead of "The word squash in line 22 means," the item in the experimental form might read "The word squash in line 27 means."

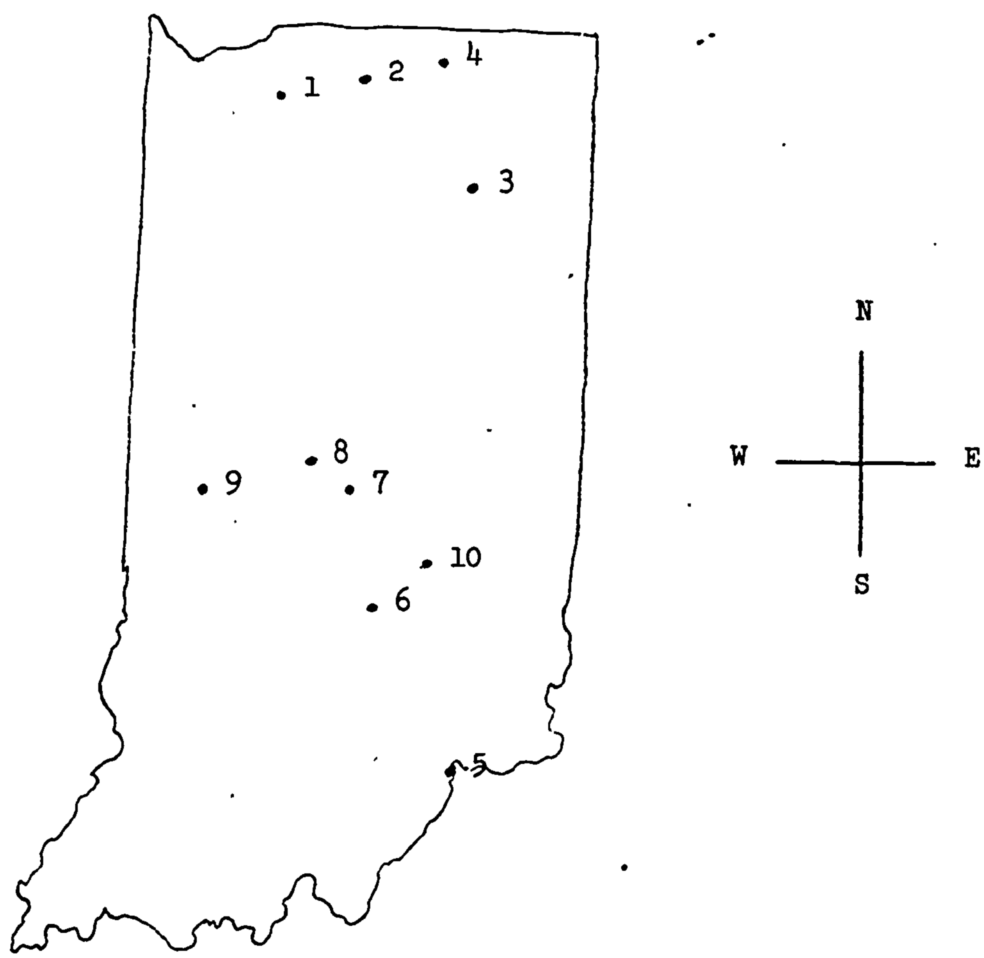
Subjects

An attempt was made to secure a sample of 4th, 5th and 6th graders which was not atypical in any specific sense. For this reason cooperating officials of the Indiana Organization of Elementary School Principals were asked to designate ten school systems (and a few back-up systems) which together would be representative of the school population of the State of Indiana. The author recognizes that this procedure does not result in the kind of representativeness associated with random sampling. However, administrative and logistical barriers to doing research in school systems selected randomly from a pool of systems are so large as to result eventually in all kinds of concessions which tend to invalidate the original purity of the sampling plan. Secondly, the selection of Ss at this stage does not involve the creation of comparison groups which require random sampling for the purpose of guaranteeing the internal validity of the research. Rather, to the degree that the actual sample used in this study is atypical of any specific population, the results will merely lack in generalizability to that population.

Figure 1 (page 13) contains a map of Indiana, and an indication of which cities provided subjects for this study. Of the 10 school systems originally invited to participate, only one declined because of involvement in another measurement oriented research project. Figure 2 shows a

Figure 1

State of Indiana



1. Valparaiso
2. LaPorte
3. Warsaw
4. Elkhart
5. Madison
6. Columbus
7. Indianapolis
8. Lebanon
9. Crawfordsville
10. Shelbyville

listing of the systems in the final sample, the number of schools in each system and the total number of students.

Figure 2

School Systems, Number of Schools, and Students Per System

	Schools	Students
1. Valparaiso Community Schools Valparaiso, Indiana	7	928
2. LaPorte Community School Corporation LaPorte, Indiana	13	1716
3. Warsaw Community School Corporation Warsaw, Indiana	9	1227
4. Elkhart Community School Corporation Elkhart, Indiana	4	822
5. Madison Consolidated Schools Madison, Indiana	7	1045
6. Bartholomew Consolidated School Corporation Columbus, Indiana	7	1334
7. Metropolitan School District of Perry Township Indianapolis, Indiana	3	720
8. Lebanon Community School Corporation Indianapolis, Indiana	4	867
9. Crawfordsville Community School Corporation Crawfordsville, Indiana	5	693
10. Shelby Eastern Schools Shelbyville, Indiana	2	99

The original design of the study called for 300 students per grade (4-6) per school system. This quota could not be met by all systems involved. In addition, the larger systems welcomed testing of as many students as were available in the cooperating schools rather than leaving some classrooms out. The resulting distribution of subjects across the various systems, in fact, offsets the apparent overrepresentation of systems in rural communities somewhat, since the few school systems contributing the most subjects are situated in the more industrial northern region of Indiana.

Table 1 gives the number of students per school system, per grade, per test and per test condition. Additional comments on this table will be provided in the next section of this report.

It may need mentioning that all students present on the day of testing in a particular school or classroom were included in the study. The only exception to this is some 25 children (in a sample of over 9,000) who did not participate because the teacher advised against it on the basis of over-anxiety or extreme inability to read.

Procedure

The administration of the tests took place during the latter half of February, March and the first half of April, 1972.

A team consisting of the author and three to five graduate research assistants administered the tests. To insure uniformity of test administration, the assistants were all trained in the procedure followed in administering the tests in order to standardize the procedures as much as possible.

Table 1

Number of Ss in Each School System Who Were Administered Tests Under Passage (P) and No Passage (NP) Conditions.

Grade 4

Test		1	2	3	4	5	6	7	8	9	10
1	NP	48	77	70	42	29	60	25	34	27	--
1	P	26	26	90	24	8	16	16	34	15	2
2	NP	32	76	78	34	38	69	17	43	27	4
2	P	25	28	33	14	32	27	16	4	13	--
3	NP	32	78	32	51	66	47	19	60	35	4
3	P	23	28	30	15	31	31	--	38	13	4
4	NP	56	86	28	24	72	50	52	22	27	7
4	P	21	40	20	16	33	20	24	18	17	2
5	NP	--	--	--	--	--	--	--	--	--	--
5	P	--	--	--	--	--	--	--	--	--	--
6	NP	27	50	79	51	60	49	36	34	32	--
6	P	--	85	--	17	21	31	20	12	13	11
Totals of Schools Tested		290	574	460	288	390	400	225	299	206	34

Grade 5

Test		1	2	3	4	5	6	7	8	9	10
1	NP	22	82	63	66	20	68	39	32	34	--
1	P	29	28	29	22	21	28	21	20	15	--
2	NP	26	75	65	70	19	68	34	37	33	10
2	P	33	21	23	22	19	32	--	38	14	2
3	NP	30	85	62	26	58	67	20	47	29	--
3	P	31	28	24	12	30	31	25	14	15	3
4	NP	--	--	--	--	--	--	--	--	--	--
4	P	--	--	--	--	--	--	--	--	--	--
5	NP	62	86	18	17	87	55	41	35	44	--
5	P	13	84	11	--	15	36	23	14	12	--
6	NP	22	79	60	33	54	69	41	37	33	8
6	P	61	22	1	16	29	43	9	22	15	--
Totals of Schools Tested		329	590	356	284	348	497	253	296	244	23

Grade 6

Test		1	2	3	4	5	6	7	8	9	10
1	NP	55	76	52	27	25	68	42	35	30	11
1	P	25	25	8	9	32	29	15	4	16	3
2	NP	29	76	60	71	[illegible]	52	30	46	34	--
2	P	25	29	39	7	20	30	--	34	15	2
3	NP	21	69	91	44	27	63	40	26	33	12
3	P	21	29	22	15	17	24	31	7	17	4
4	NP	--	--	--	--	--	--	--	--	--	--
4	P	--	--	--	--	--	--	--	--	--	--
5	NP	56	78	47	25	69	37	24	34	43	3
5	P	--	43	29	16	42	20	44	--	12	3
6	NP	21	85	58	36	51	69	6	59	27	2
6	P	56	42	5	--	8	45	10	27	16	2
Totals of Schools Tested		309	552	411	250	307	437	242	272	243	42

Key for School System

1 - Valparaiso
2 - LaPorte
3 - Warsaw
4 - Elkhart
5 - Madison
6 - Columbus
7 - Perry
8 - Lebanon
9 - Crawfordsville
10 - Shelbyville

Since the purpose of the study was to obtain passage dependency indices on all items in the tests used, no time limits were enforced. The standard directions used for the purpose of this study included:

a. Mentioning of the fact that the tests were administered for the purpose of getting information on the tests and not on the children.

b. The statement that the results would not appear on grade cards, or be reported to the teachers.

c. A plea for cooperation.

d. An explanation of how to use the test booklet with the passage booklets.

e. The encouragement that "many questions can be answered without reading the stories." (Only for the children under the No-Passage Condition).

f. The announcement that there would be plenty of time.

Depending on the test, the administration of the tests under the Passage condition lasted typically from 45-60 minutes. Under the No-Passage condition, about 20-45 minutes were needed.

Cooperating schools were given the option to have their students tested in large groups in cafeterias, etc. or in classrooms. This decision was made on the basis of the results of Ingle and DeAmico (1969) who found no effect of physical conditions on standardized achievement test scores. The conditions contrasted in their study were "relatively poor physical conditions in an auditorium" and "relatively adequate physical conditions in regular classrooms." The principals of the schools in the present study, in general, preferred testing in classrooms. Thus, only approximately ten percent of all test administrations took place in an auditorium or cafeteria.

As indicated, two thirds of Ss took the tests without the passages. Assignment to the Passage condition (P) or Non-passage condition (NP) was done with the classroom as the unit. The argument for this decision was that the confusion resulting from the differences in time needed for completion of the task and the necessity for two sets of directions if both N and NP students would be present in one classroom would outweigh any advantages due to using the student as the unit of assignment.

Responses were recorded on machine scoreable answer sheets. Great care was taken to insure that students knew how to use these. Infrequent problems in this respect were detected early, since, routinely, both the classroom teacher and the E monitored during the first ten minutes of the test administration.

RESULTS

1. Results Combined Over Grades

Table 2 (See page 19) summarizes the scores on all tests across the three grades. This table invites a few comments. First of all, it is clear that deviation from the publishers' standard test administration procedure had little effect on the reliability of the measurements. With the exception of Test 6 (a subset of items of the ITBS) all reliability coefficients under the P-condition are equal to or above .90. The fact that under the NP-condition the KR-20's are lower is not surprising. After all, in this condition the task is to guess at the answer. What *is* surprising is that the reliabilities remain as high as they do. This in

Table 2

Means, Standard Deviations, KR-20 Coefficients and Standard Errors of Measurement for the P and NP Conditions Across Grades

Condition	Test	k	$\bar{X}$	S.D.	KR-20	SE_m
Passage	1	75	45.96	15.8	.96	3.3
	2	42	26.66	8.1	.90	2.6
	3	60	37.17	12.3	.93	3.2
	4	45	29.54	9.4	.92	2.7
	5	45	28.82	8.7	.90	2.7
	6	42	27.03	7.7	.88	2.7
Non-Passage	1	75	29.36	6.7	.67	3.9
	2	42	14.36	4.1	.51	2.9
	3	60	22.17	6.3	.70	3.5
	4	45	22.27	6.7	.81	3.0
	5	45	20.27	5.0	.65	3.0
	6	42	19.29	5.0	.68	2.9

itself is an indication that the behavior measured is not a random selecting of any of four multiple-choice options.

The mean scores under the P-condition are in the expected range, typically some 60% of the highest possible score. The decision to allow more time than the test manuals specify, however, makes it impossible to interpret the scores of the P-students in terms of the norms provided in the manuals.

From Table 2 it can already be seen that none of the tests produces mean scores under the NP-condition close to what one would expect on the basis of chance only. For all tests, with the exception of Test 2, this chance score equals r/4, where r = the number of items. Test 2 contains a few five choice items; the chance score for this test equals 10.10. Table 3 details the extent to which the scores under the passage con-

Table 3

Means Under the NP-condition Expressed as (1) Percentages of the Number of Items in the Test and (2) As Percentages of the Means Obtained Under the P Condition

Test	$\bar{X}_P$	$\bar{X}_{NP}$	$\bar{X}_{NP}$ as % of Number of Items	Chance Score (%)	$\bar{X}_{NP}$ as % of X_P
1	45.96	29.36	39	25	64
2	26.66	14.36	34	24	54
3	37.17	22.17	37	25	60
4	29.54	22.27	50	25	75
5	28.82	20.27	45	25	70
6	27.03	19.29	46	25	71

dition exceeded chance scores. The entries in the cells can be contrasted directly with those in column 5, representing chance scores. It is clear that none of the tests even approximates the chance score under the NP-condition. Tests 4, 5 and 6, in particular, show a high degree of passage independency. The fourth graders to whom Test 4 was administered managed to answer correctly 50% of the items even though they never read the passage upon which the items were based. Tests 5 and 6 fare little better and even Tests 1, 2 and 3 result in "guessing" scores which are far above the level of statistical chance.

The entries in the last column are even more startling. Of the six tests, three allow a student who does not have the passages to obtain a score as high as 70% of what a student with the passages would get. On the average, for these tests, not reading the passage results in a loss of performance less than 30%. Tests 1, 2 and 3 present only a slightly more reassuring picture. It may be noted that if one takes 60% of the number of items as a typical mean score for multiple choice tests, the expected chance score of 25% represents approximately 40% of the score obtained under the P-condition (60%). Again, Test 2 shows up more favorably than the other tests.

2. Results by Grade

Tables 4 through 13 (See pages 22-26) contain the results presented above broken down by grade. Passage dependency of items is not a static characteristic. It varies with the test user. This is a potential problem if one particular test form is used for a number of grade levels. The nature and the extent of the problem are illustrated below.

Table 4 needs little commentary except to note that there is an increase of the means across grades in both the P and the NP-condition. Whereas this is not surprising in the former case, the increases under the NP-condition are of some interest. The data indicate that a particular item may be sufficiently passage dependent at the lower level of the grade range for which the test was intended but insufficiently passage dependent at the higher levels.

Table 4

Test 1 - Means, Standard Deviations, KR-20 Reliability Coefficients and Standard Errors of Measurement for P and NP-Conditions. (Number of Items = 75; Chance Score = 18.75.) Results by Grade.

Condition	Grade	$\bar{X}$	S.D.	KR-20	SE_m
P	4	39.69	16.5	.96	3.3
	5	47.58	14.0	.95	3.2
	6	53.46	13.0	.94	3.1
NP	4	28.74	6.4	.64	3.9
	5	28.86	6.9	.69	3.8
	6	30.53	6.8	.67	3.9

Table 5

Test 1 - Means Under the NP-Condition Expressed (1) as Percentages of the Number of Items in the Test and (2) as Percentages of the Mean Under the P-Condition. Results by Grade.

Grade	Means as % of Total Number of Items	Mean as % of Means Under Passage Condition
4	38	72
5	38	61
6	41	57

Table 6

Test 2 - Means, Standard Deviations, KR-20 Reliability Coefficients and Standard Errors of Measurement for P and NP-Conditions. (Number of Items = 42; Chance Score = 10.10) Results by Grade.

Condition	Grade	$\bar{X}$	S.D.	KR-20	SE_m
P	4	23.05	7.18	.85	2.77
	5	28.04	7.47	.88	2.55
	6	28.92	8.25	.91	2.47
NP	4	13.34	3.83	.44	2.88
	5	14.35	4.10	.51	2.87
	6	15.42	4.24	.54	2.88

Table 7

Test 2 - Means Under the NP-Condition Expressed (1) as Percentages of the Number of Items in the Test and (2) as Percentages of the Mean Under the P-Condition. Results by Grade.

Grade	Means as % of Total Number of Items	Mean as % of Means Under Passage Condition
4	32	58
5	34	51
6	37	53

Table 8

Test 3 - Means, Standard Deviations, KR-20 Reliability Coefficients and Standard Errors of Measurement for P and NP-Conditions. Number of Items = 60; Chance Score = 15.00) Results by Grade.

Condition	Grade	$\bar{X}$	S.D.	KR-20	SE_m
	4	31.72	11.49	.92	3.29
P	5	38.76	12.03	.93	3.14
	6	41.25	11.51	.93	3.05
	4	19.97	5.29	.58	3.41
NP	5	21.91	6.06	.67	3.45
	6	24.61	5.50	.72	3.46

Table 9

Test 3 - Means Under the NP-Condition Expressed (1) as Percentages of the Number of Items in the Test and (2) as Percentages of the Mean Under the P-Condition. Results by Grade.

Grade	Means as % of Total Number of Items	Mean as % of Means Under Passage Condition
4	33	63
5	37	57
6	41	60

Table 10

Test 5 - Means, Standard Deviations, KR-20 Reliability Coefficients and Standard Errors of Measurement for P and NP-Conditions. (Number of Items = 45; Chance Score = 11.25) Results by Grade.

Condition	Grade	$\bar{X}$	S.D.	KR-20	SE_m
P	4	29.54	9.41	.92	2.67
	5	27.19	8.77	.90	2.80
	6	30.43	8.32	.90	2.65
NP	4	22.27	6.70	.81	2.96
	5	19.16	4.87	.62	2.99
	6	21.45	4.86	.63	2.95

Table 11

Tests 4 and 5 - Means Under the NP-Condition Expressed (1) as Percentages of the Number of Items in the Test and (2) as Percentages of the Mean Under the P-Condition. Results by Grade.

Grade	Test	Means as % of Total Number of Items	Mean as % of Means Under Passage Condition
4	4	50	75
5	5	43	70
6	5	48	70

Table 12

Test 6 - Means, Standard Deviations, KR-20 Reliability Coefficients and Standard Errors of Measurement for P and NP-Conditions. (Number of Items = 42; Chance Score = 10.50) Results by Grade.

Condition	Grade	$\bar{X}$	S.D.	KR-20	SE_m
P	4	23.75	7.92	.88	2.74
	5	27.82	7.55	.88	2.61
	6	29.47	6.56	.85	2.58
NP	4	17.85	4.82	.65	2.85
	5	19.34	4.95	.67	2.85
	6	20.67	4.96	.68	2.81

Table 13

Test 6 - Means Under the NP-Condition Expressed (1) as Percentages of the Number of Items in the Test and (2) as Percentages of the Mean Under the P-Condition. Results by Grade.

Grade	Mean as % of Total Number of Items	Mean as % of Means Under Passage Condition
4	42	75
5	46	70
6	49	70

The entries in the second column of Table 5 indicate the absolute increase in passage independency of the test items as the student becomes more sophisticated. The entries in the third column are of particular interest. They index the performance under the NP-condition relative to that under the P-condition for a particular grade group. The decrease in the percentages indicates that the means under the P-condition increase faster than those under the NP-condition. In summary, the data in Table 5 may be interpreted as follows. Whereas a student in the 6th grade who fails to read the passages (or some of them) can get more answers right than a 4th grader in the same position, the score of the 6th grader relative to those of his peers will be lower than the score of the 4th grade student when compared to the scores of other 4th graders.

Tables 6-13 show patterns similar to the ones discussed for Test 1 for the remaining tests. It may be pointed out, however, that the increase in relative passage dependency does, in general, not hold for grades 5 and 6. This can be seen from inspection of the entries in the last column of Tables 7, 9, 11 and 13. Two of the tests, 2 and 3, even show a reversal; it is too small to be of any importance, however (Tables 7 and 9). Care must be taken not to misinterpret the data in Tables 10 and 11, where the combined results for Tests 4 and 5 are reported. Of special interest is the very high reliability coefficient for test 4 under the NP-condition. Table 11 reveals that Ss in this condition obtained a mean score as high at 75% of the mean score under the P-condition. The items on this particular test allowed the NP subjects to employ a highly reliable and effective response strategy.

The entries in the column headed "Means as % of the Number of Items" of Tables 7, 9 and 13 show the same absolute increase in passage independency as a function of increased sophistication of the respondents, as was noted for Test 1. For Test 5, the comparison by grade is limited to grades 5 and 6 (Table 11).

Items With Higher than Chance Scores

There are two ways in which the mean score of a group of subjects under the NP-condition can be higher than r/k, when r = the number of items and k = the average number of options per item. First, there may be a relatively small group of very easy items. Secondly, there may be a large group of moderately easy items, all of which, however, have a probability of being passed larger than $1/k$.

Additional light on the mean scores reported in Tables 4-13 is supplied by the information in Table 14 (See page 29) and Table 15 (See page 30). The 1% and 5% upper confidence limits (one-sided) were calculated for each group of respondents to each test. The basis for the calculations is the binomial distribution where $p = 1/k$ and $C.L. = 1-p$. The limits were computed around the quantity $1/k$, in most cases equal to .25. An item is said to have a passage independency larger than $1/k$, if the observed item difficulty exceeded the upper confidence limit. Table 14 shows, for each test, the number of items per test with a passage independency larger than $1/k$. In Table 15 the same information is expressed as percentages of the number of items per test.

Table 15 in particular points up a few interesting characteristics of the tests analyzed. First of all, it becomes clear that, in general

Table 14

Number of Items Per Test with a Difficulty Under the NP-Condition
Higher than 1/k, Where k = the Number of Options Per Item.
Items Included Had an Observed Difficulty Exceeding
One-sided Upper Confidence Limits Around (1/k).
The Main Entries are Based on 5% C.L.'s;
the Entries in Parentheses are Based
on 1% C.L.'s.

Test	# of Items	Grade 4	Grade 5	Grade 6
1	75	49 (46)	48 (45)	51 (49)
2	42	25 (22)	26 (25)	25 (24)
3	60	28 (25)	32 (29)	40 (37)
4	45	37 (37)	— (—)	— (—)
5	45	— (—)	32 (30)	33 (32)
6	42	31 (30)	34 (33)	33 (33)

Table 15

Percentage of Items Per Test with a Difficulty Under the NP-Condition Higher than 1/k, Where k = the Number of Options Per Item. Items Included Had an Observed Difficulty Exceeding One-sided Upper Confidence Limits Around (1/k). The Main Entries are Based on 5% C.L.'s; the Entries in Parenthesis are Based on 1% C.L.'s.

Test	Grade 4	Grade 5	Grade 6
1	65 (61)	64 (60)	68 (65)
2	60 (52)	62 (60)	60 (57)
3	47 (42)	53 (48)	67 (62)
4	82 (82)	— (—)	— (—)
5	— (—)	71 (67)	73 (71
6	69 (67)	76 (73)	73 (73)

tests with the highest mean scores (relative to the number of items) under the NP-condition also have the highest percentage of items with a passage independency index larger than 1/k. This can be seen by contrasting Table 15 with the entries in Tables 5, 7, 9, 11 and 13. Test 4, for instance, had an NP mean score which was 50% of the total number of items in this test (Table 11). From Table 15 it can be seen that 82% of those items had a NP-difficulty index exceeding the 1% upper confidence limits around 1/k, where k = the average number of options per item. A comparison of the results for Tests 2 and 3, however, shows an exception to this general finding. From Tables 7 and 9 it can be seen that Test 2 has NP mean scores which are, if expressed as a proportion of the total number of items in the test, smaller than the NP mean scores for Test 3. Yet, for grades 4 and 5 the percentage of items with a passage independency index larger than 1/k is higher for Test 2 than it is for Test 3. Test 3, relative to Test 2, is a test where a high NP mean score is obtained with relatively few passage independent items. Tables A1 - A6 contain the NP-difficulties for each item and a designation in regard to whether or not these exceeded 5% and 1% C.L.'s around 1/k, where k is the number of options per item.

Passage Dependency Indices

The data presented above makes clear that generally quite a few items allow respondents to answer correctly when they have not read the material upon which the items purportedly were based. The degree to which an item requires reading of the passage has been referred to as that item's passage dependency. The term passage independency has been used to indicate the (relative) lack of passage dependency. When the question arises as to how to index numerically passage dependency of an item, the most logical course of action seems to be to obtain an estimate of the proportion of respondents that can answer the item under the NP-condition. Thus,

Passage Dependency Index 1 = The proportion of correct responses under the NP condition

Thus, the lower PDI_1 is, the more passage dependent the item for which it was calculated. Since the concept of "the lower the better" may be slightly confusing, it is better to calculate:

$$PDI_2 = 1 - PDI_1$$

This index will increase as passage dependency increases.

Theoretically, PDI_1 can take any value between 1/k and 1.00, where k = the number of options per item. For most multiple-choice tests the range of PDI_1 would be .25 - 1.00. However, certain items have characteristics which lower the probability of choosing the correct response under the NP-condition. Actual values of PDI_1 lower than .25 may therefore be observed. Conversely, PDI_2 values higher than the theoretical .75 do occur frequently.

Table 16 (See page 33) contains the PDI_2 values for the tests.

In Tables B1 - B2 (See Appendix B) the PDI_1 (d_{NP}) and PDI_2 values for all the items have been listed.

There is a problem in interpreting PDI_1 and PDI_2, however. Consider the following statistics for two items. Item 1 has a difficulty d_P under the P condition of .35 and a difficulty d_{NP} under the NP condition of

Table 16

PDI_2 Values for Tests 1-6;
By Grade

Test	PDI_2			
	Grade			
	4	5	6	7
1	62	62	59	61
2	68	66	63	66
3	67	63	59	63
4	-	-	-	-
5	-	57	52	55
6	58	54	51	54

.35 also. Item 2 has a d_P equal to .75 and, like item 1, a d_{NP} value of .35. For both items PDI_2 = .65. However, while it seems difficult to use item 1 for observing any behavior controlled by the passage, item 2 can be used to this end. After all, at least 40% of correct responses found their sources in the passage. (The question of correctly guessing is left aside for the moment.) For this reason, while PDI_2 provides some information about an item's passage dependency, it does not tell the whole story. Tuinman (1970) proposed the ratio of d_{NP}/d_P as a better index of the degree to which a question can be used efficiently to measure responses based on reading the passage. Thus,

$$E_1 = d_{NP} / d_P$$

For item 1 in the example above $E_1 = 1.00$ and for item 2 $E_1 = .50$.
For convenience of interpretation E_2 is proposed:

$$E_2 = 1 - d_{NP}/d_P$$

Table 17 contains the average d_{NP} and d_P values for the six tests and the resulting E_2 values.

Table 17

Average Difficulties Under P and NP Conditions and E_2 Values for Six Tests; Combined Across Grades

Test	d_P	d_{NP}	E_2
1	.61	.39	.36
2	.63	.34	.46
3	.62	.37	.40
4	.66	.50	.24
5	.64	.45	.30
6	.64	.46	.28

Since the six tests have comparable difficulties under the P-condition, E_2 is of importance in particular for comparison of items of tests with different d_{NP} values. Tables B1 - B6 contain the d_P, d_{NP}, PDI_2 and E_2 values for all the items.

The indices proposed above are all very simple and make no special assumptions. In Appendix C a number of other indices of passage dependency are discussed. Tables C1 - C6 contain the values of these statistics for all items.

Discussion

This report has a two fold purpose. First, it intends to highlight the problem of passage dependency of reading comprehension items; secondly, it is meant to be a working document for those who desire to do further analyses on the items included in the tests used. For this purpose extensive data tables have been included in Appendices A and B.

From the data presented above, a number of major conclusions can be drawn. First of all, it appears that commercially marketed tests of reading comprehension vary considerably in the degree to which their items are passage dependent. This points up the need to consider passage dependency when choosing among various tests. Everything else being equal, the test with the most items with the highest degree of passage dependency offers the largest guarantee against invalidity due to responding to items without prior reading of the passage on which the item is based. A caution is in place, however; passage dependency may be purchased at a price that the test consumer is unwilling to pay. The Weaver, Bickley and Ford (1969) study, for instance, indicates that, generally, inference items are more passage dependent than factual items. In addition to considering passage dependency, the consumer must satisfy himself in regard to the content validity of the test under consideration. Secondly, it becomes obvious from the data presented above, that none of the five tests approaches passage dependency close to optimal limits.

This is not so surprising in view of the fact that it is extremely difficult to construct highly passage dependent items, even if the passages contain highly imaginary materials (Tuinman, 1970). Thirdly, as expected, (Tuinman, 1971), the degree to which items are passage dependent is a function of the age, c. q. educational sophistication of the child that takes the test. Tests in this study, designed for grades 4-6, showed a consistent decrease in passage dependency from fourth graders to sixth graders. This fact must be kept in mind by the test user who decides to select a test with a wider grade range.

References

Bickley, A. C., Weaver, W., & Ford, Fraughton. Information removed from multiple-choice item responses by selected grammatical categories. Psychological Reports, 1968, 23, 613-614.

Bloomer, R. H., & Heitzman, A. J. Pre-testing and the efficiency of paragraph reading. Journal of Reading, 1965, 8, 219-223.

Buros, O. K. The sixth mental measurements yearbook. Highland Park, New Jersey: The Gryhon Press, 1965.

Carney, J. D., & Scheer, R. K. Fundamentals of logic. New York: The Macmillan Company, 1964, 207.

Christensen, C. M., & Stordahl, K. E. The effect of organizational aids on comprehension and retention. Journal of Educational Psychology, 1955, 46, 65-74.

Davis, F. B. What do reading tests really measure? The English Journal, 1944, 33, 180-187.

Davis, F. B. Research in comprehension in reading. Reading Research Quarterly, 1968, 3, 499-545.

Eurich, A. C. A method for measuring retention in reading. Journal of Educational Research, 1931, 24, 202-208.

Farr, R., & Smith, C. B. The effects of test item validity on total test reliability and validity. In G. Schick and M. M. May (Eds.), Reading: Process and pedagogy. Nineteenth yearbook of the national reading conference, Vol. 1, Milwaukee, Wisconsin, 1970, 122-134.

Ingle, Robert B., & DeAmico, Gerald. The effect of physical conditions of the test room on standardized achievement test scores. Journal of Educational Measurement, 1969, 6, 237-240.

Mitchell, R. W. A comparison of children's responses to an original and experimental form of subtests GS and ND of the Gates Basic Reading Tests. Unpublished Doctoral Dissertation, University of Minnesota, Dissertation Abstracts, 1967, 9704A.

Preston, R. C. Ability of students to identify correct responses before reading. Journal of Educational Research, 1964, 58, 181-183.

Samuels, S. Jay Effect of word associations on reading speed, recall, and guessing behavior on tests. Journal of Educational Psychology, 1968, 59, 12-15.

Tuinman, J. J. Selected aspects of the assessment of the acquisition of information from reading passages. Unpublished Doctoral Dissertation, University of Georgia, 1970.

Tuinman, J. Asking passage dependent reading questions. Journal of Reading, 1971, 14 (5), 289-292, 336.

Tuinman, J. Jaap. Children's willingness to skip reading passages when taking reading comprehension tests. The Southern Journal of Educational Research, 1972, 6, 1-13. (a)

Tuinman, J. Jaap. Inspection of reading comprehension passages as a function of passage dependency of test items. Research Report of the Institute for Child Study, Indiana University, Bloomington, Indiana, 1972. (b)

Weaver, W., & Bickley, A. C. Sources of information for responses to reading test items. APA Proceedings, 75th Annual Convention, 1967, 293-294.

Weaver, W. W., Bickley, A. C., & Ford, F. A cross-validation study of the relationship of reading test items to their relevant paragraphs. Perceptual and Motor Skills, 1969, 29, 11-14.

APPENDIX A

Tables (A1-A6) Per Test for Item Difficulties
Under the NP-Conditions, Per Grade.

Table A1

Test 1 - Item Difficulties Under the NP-Conditions, Per Grade. No Asterisk Indicates that the Item Difficulty Exceeded the 1% Upper Confidence Limit Around (1/k). One Asterisk Indicates that the Difficulty was Higher than the 5% C.L. but not Higher than the 1% C.L. Two Asterisks Indicate that the Difficulty did not Exceed the 5% C.L. (Decimal Points have been Deleted.)

Item	Grade 4	Grade 5	Grade 6	Item	Grade 4	Grade 5	Grade 6
1	82	80	83	39	39	40	42
2	31	30*	31	40	19**	19**	17**
3	49	50	52	41	29**	32	32
4	89	82	83	42	30*	27**	36
5	40	46	49	43	26**	23**	29**
6	77	80	79	44	15**	16**	14**
7	37	42	47	45	15**	14**	14**
8	71	66	74	46	31	37	39
9	50	46	46	47	24**	26**	27**
10	14**	18**	18**	48	36	37	31
11	66	66	59	49	37	47	47
12	60	63	68	50	24**	18**	23**
13	69	67	66	51	50	50	52
14	78	81	84	52	39	42	51
15	51	56	61	53	44	44**	46
16	60	62	69	54	28**	24**	23**
17	60	59	60	55	33	34	35
18	09**	06**	08**	56	65	66	67
19	45	42	43	57	60	64**	67
20	32	30	34	58	19**	16*	20**
21	54	51	54	59	21**	29**	29*
22	39	31	35	60	28**	26**	24**
23	07**	06**	06**	61	20**	21**	20**
24	50	55	59	62	46	48	53
25	12**	13**	14**	63	41	38	43
26	28**	24**	28**	64	21**	22**	23**
27	29*	25**	34	65	13**	20**	19**
28	50	52	56	66	19**	16**	28**
29	19**	21**	21**	67	22**	20**	27**
30	45	42	40	68	34	40	40
31	58	55	57	69	49	47	51
32	48	41	46	70	15**	12**	11**
33	56	56	59	71	19**	17**	14**
34	29*	41	50	72	42	48	52
35	44	50	54	73	29*	27**	30*
36	22**	22**	24**	74	22**	22**	22**
37	28**	30*	31	75	32	28**	28**
38	45	45	52				

Table A2

Test 2 - Item Difficulties Under the NP-Conditions, Per Grade. No Asterisk Indicates that the Item Difficulty Exceeded the 1% Upper Confidence Limit Around (1/k). One Asterisk Indicates that the Difficulty was Higher than the 5% C.L. but not Higher than the 1% C.L. Two Asterisks Indicate that the Difficulty did not Exceed the 5% C.L. (Decimal Points have been Deleted.)

Item	Grade 4	Grade 5	Grade 6	Item	Grade 4	Grade 5	Grade 6
1	13**	14**	13**	22	31	24**	22**
2	11**	09**	08**	23	29*	31	28**
3	36	38	40	24	65	71	71
4	29*	36	30*	25	55	64	72
5	23**	22**	17**	26	26**	18**	14**
6	20**	17**	18**	27	65	75	77
7	32	31	38	28	36	43	39
8	37	41	47	29	52	55	51
9	12**	18**	18**	30	48	53	62
10	23**	21**	31	31	29*	24**	25**
11	34	36	44	32	32	30*	26**
12	32	43	42	33	36	36	42
13	36	31	39	34	39	48	61
14	27**	36	48	35	24**	31	32
15	35	45	50	36	27**	24**	19**
16	39	43	50	37	22**	20**	27**
17	51	44	52	38	14**	16**	17**
18	39	49	53	39	16**	16**	19**
19	42	46	50	40	23**	21**	21**
20	23**	29*	41	41	23**	20**	23**
21	32	43	43	42	19**	20**	23**

Table A3

Test 3 - Item Difficulties Under the NP-Conditions, Per Grade. No Asterisk Indicates that the Item Difficulty Exceeded the 1% Upper Confidence Limit Around (1/k). One Asterisk Indicates that the Difficulty was Higher than the 5% C.L. but not Higher than the 1% C.L. Two Asterisks Indicate that the Difficulty did not Exceed the 5% C.L. (Decimal Points have been Deleted.)

Item	Grade 4	Grade 5	Grade 6	Item	Grade 4	Grade 5	Grade 6
1	57	61	67	31	15**	19**	15**
2	80	80	85	32	23**	29*	35
3	61	56	60	33	26**	29*	34
4	64	59	69	34	47	59	67
5	11**	14**	18**	35	17**	18**	22**
6	50	48	56	36	22**	33	39
7	48	56	60	37	25**	23**	24**
8	37	41	52	38	22**	26**	28**
9	17**	23**	28**	39	25**	36	37
10	48	47	50	40	30*	33	42
11	24**	25**	20**	41	24**	22**	21**
12	34	41	48	42	31	37	46
13	25**	28**	29*	43	30*	30*	31
14	72	75	80	44	29*	28**	28**
15	19**	30*	40	45	21**	22**	23**
16	37	54	59	46	51	57	67
17	54	61	62	47	26**	28**	29*
18	32	27**	36	48	20**	21**	21**
19	23**	29*	35	49	27**	23**	25**
20	26**	28**	28**	50	28**	28**	34
21	42	48	47	51	38	40	51
22	46	48	55	52	19**	21**	26**
23	66	71	79	53	34	39	41
24	54	65	82	54	21**	24**	27**
25	20**	20**	23**	55	22**	27**	30*
26	28**	26**	30*	56	29*	26**	37
27	47	59	74	57	23**	25**	24**
28	14**	17**	20**	58	23**	25**	27**
29	45	53	58	59	26**	28**	30*
30	19**	25**	29*	60	25**	19**	23**

Table A4

Test 4 - Item Difficulties Under the NP-Conditions, Per Grade. No Asterisk Indicates that the Item Difficulty Exceeded the 1% Upper Confidence Limit Around (1/k). One Asterisk Indicates that the Difficulty was Higher than the 5% C.L. but no Higher than the 1% C.L. Two Asterisks Indicate that the Difficulty did not Exceed the 5% C.L. (Decimal Points have been Deleted.)

Item	Grade 4	Item	Grade 4
1	51	23	18**
2	39	24	72
3	79	25	70
4	81	26	46
5	48	27	46
6	61	28	46
7	47	29	16**
8	81	30	57
9	17**	31	25**
10	74	32	83
11	25**	33	63
12	71	34	41
13	37	35	14**
14	53	36	50
15	50	37	45
16	71	38	68
17	49	39	18**
18	66	40	59
19	21**	41	43
20	38	42	57
21	41	43	75
22	39	44	38
		45	38

Table A5

Test 5 - Item Difficulties Under the NP-Conditions, Per Grade. No Asterisk Indicates that the Item Difficulty Exceeded the 1% Upper Confidence Limit Around (1/k). One Asterisk Indicates that the Difficulty was Higher than the 5% C.L. but no Higher than the 1% C.L. Two Asterisks Indicate that the Difficulty did not Exceed the 5% C.L. (Decimal Points have been Deleted.)

Item	Grade 5	Grade 6	Item	Grade 5	Grade 6
1	78	83	23	35	40
2	68	84	24	53	53
3	75	86	25	31	34
4	47	56	26	28**	35
5	41	40	27	42	57
6	38	44	28	76	80
7	64	75	29	55	66
8	65	72	30	23**	30
9	63	67	31	20**	18**
10	55	68	32	48	47
11	57	62	33	31	39
12	18**	30*	34	14**	08**
13	29*	28	35	23**	28**
14	78	88	36	21**	23**
15	63	60	37	20**	25**
16	78	80	38	39	45
17	11**	12**	39	32	37
18	19**	19**	40	33	42
19	47	59	41	22**	22**
20	35	45	42	24**	21**
21	68	72	43	24**	26**
22	51	55	44	51	60
			45	24**	24**

Table A6

Test 6 - Item Difficulties Under the NP-Conditions, Per Grade. No Asterisk Indicates that the Item Difficulty Exceeded the 1% Upper Confidence Limit Around (1/k). One Asterisk Indicates that the Difficulty was Higher than the 5% C.L. but not Higher than the 1% C.L. Two Asterisks Indicate that the Difficulty did not Exceed the 5% C.L. (Decimal Points have been Deleted.)

Item	Grade 4	Grade 5	Grade 6	Item	Grade 4	Grade 5	Grade 6
1	07**	09**	05**	22	64	76	78
2	27**	31	34	23	41	42	43
3	79	81	86	24	45	50	65
4	88	89	91	25	24**	25**	24**
5	02**	03**	03**	26	54	56	61
6	80	81	87	27	50	50	61
7	64	65	72	28	23**	33	48
8	50	47	47	29	19**	21**	26**
9	79	81	82	30	62	73	82
10	45	50	46	31	16**	22**	23**
11	41	43	44	32	38	52	58
12	55	65	70	33	36	32	37
13	62	64	70	34	16**	19**	19**
14	38	30*	27**	35	49	58	52
15	26**	36	39	36	41	40	[illegible]
16	49	52	60	37	13**	14**	17**
17	52	52	53	38	45	43	42
18	37	44	51	39	29*	36	40
19	37	46	45	40	35	36	46
20	18**	19**	18**	41	43	49	46
21	78	86	91	42	31	33	33

APPENDIX B

For Each Test, Difficulty Coefficients (Tables B1-B6) Under P and NP Conditions (d_P, d_{NP}), Passage Dependency Index 2 ($PDI_2 = 1-d_{NP}$) and Passage Dependency Efficiency Index ($E_2 = 1-d_{NP}/d_P$).

Table B1

Test 1 - Difficulty Coefficients Under P and NP-Conditions (d_P, d_{NP}), Passage Dependency Index 2* and Passage Dependency Efficiency Index (E_2)** (Combined Across Grades; Decimal Points are Deleted.)

Item	E_2	d_P	d_{NP}	PDI_2	Item	E_2	d_P	d_{NP}	PDI_2
1	14	95	82	18	39	47	76	40	60
2	66	89	31	70	40	66	55	19	82
3	45	92	51	49	41	27	43	31	69
4	10	94	85	15	42	46	58	31	69
5	51	91	45	55	43	30	38	26	74
6	10	87	79	21	44	72	54	15	85
7	53	90	42	58	45	68	45	15	86
8	21	89	70	30	46	32	52	36	65
9	42	81	47	53	47	53	55	26	74
10	80	83	17	83	48	49	68	35	65
11	25	85	64	36	49	29	61	44	57
12	27	88	64	36	50	62	57	22	78
13	19	83	67	33	51	23	66	51	49
14	10	90	81	19	52	29	62	44	56
15	38	90	56	44	53	32	66	45	55
16	28	88	64	36	54	49	49	25	75
17	32	87	60	41	55	37	54	34	66
18	90	78	08	92	56	-08	61	66	34
19	38	69	43	57	57	-07	34	64	36
20	-3.57	08	32	68	58	45	34	18	82
21	40	88	53	47	59	30	38	26	74
22	60	87	35	65	60	34	40	26	74
23	91	70	06	94	61	39	33	20	80
24	20	68	55	45	62	-02	48	49	51
25	83	75	13	87	63	-04	39	41	59
26	58	64	27	74	64	36	37	24	76
27	61	76	30	71	65	53	37	17	83
28	31	76	53	47	66	18	15	18	82
29	72	73	21	80	67	03	24	23	77
30	31	61	42	58	68	02	38	38	62
31	34	85	56	44	69	19	41	49	51
32	41	76	45	55	70	-07	12	13	87
33	31	82	57	43	71	58	40	17	83
34	30	56	40	60	72	48	32	47	53
35	30	71	49	51	73	63	18	29	71
36	62	59	23	77	74	13	25	22	78
37	51	60	29	71	75	-1.20	13	29	71
38	38	75	47	53					

* $PDI_2 = 1 - d_{NP}$

** $E_2 = 1 - d_{NP}/d_P$

Table B2

Test 2 - Difficulty Coefficients Under P and NP Conditions (d_P, d_{NP}), Passage Dependency Index 2* and Passage Dependency Efficiency Index (E_2)**

Item	E_2	d_P	d_{NP}	PDI_2	Item	E_2	d_P	d_{NP}	PDI_2
1	86	94	14	87	22	03	26	26	75
2	90	93	10	90	23	61	18	29	71
3	49	74	38	62	24	17	84	69	31
4	61	81	31	67	25	16	76	64	36
5	66	62	21	79	26	77	85	19	81
6	74	71	19	82	27	08	78	72	28
7	59	82	34	66	28	45	71	39	61
8	48	80	42	58	29	08	58	53	47
9	81	85	16	84	30	11	49	54	46
10	70	83	25	75	31	28	37	26	74
11	45	69	38	62	32	45	54	30	70
12	42	67	39	61	33	10	43	38	62
13	18	43	35	65	34	08	53	49	51
14	56	84	37	63	35	61	75	29	71
15	27	60	43	57	36	69	76	23	77
16	40	73	44	56	37	64	64	23	77
17	38	79	49	51	38	76	65	15	85
18	34	71	47	53	39	68	55	17	83
19	21	59	46	54	40	58	52	22	78
20	30	44	31	69	41	27	30	22	78
21	04	38	39	61	42	29	29	21	80

* $PDI_2 = 1-d_{NP}$

** $E_2 = 1-d_{NP}/d_P$

Table B3

Test 3 - Difficulty Coefficients Under P and NP Conditions (d_P, d_{NP}), Passage Dependency Index 2* and Passage Dependency Efficiency Index (E_2)**

Item	E_2	d_P	d_{NP}	PDI_2	Item	E_2	d_P	d_{NP}	PDI_2
1	35	95	62	39	31	73	61	16	84
2	11	92	82	18	32	49	57	29	71
3	29	84	59	41	33	45	54	30	70
4	25	85	64	36	34	26	78	58	42
5	84	88	14	86	35	73	69	19	81
6	32	75	51	49	36	56	71	31	69
7	28	76	54	46	37	42	42	24	76
8	45	79	44	56	38	58	60	25	75
9	48	44	23	77	39	18	40	33	67
10	48	92	48	52	40	38	57	35	65
11	64	64	23	77	41	71	76	22	78
12	36	63	41	59	42	11	43	38	62
13	57	64	27	73	43	49	59	30	70
14	15	88	75	25	44	39	46	28	72
15	56	68	30	70	45	49	43	22	78
16	25	66	50	50	46	09	64	58	42
17	01	60	59	41	47	49	54	28	72
18	45	58	32	68	48	45	38	21	79
19	45	53	29	71	49	62	66	25	75
20	67	82	27	73	50	43	52	30	70
21	43	81	46	54	51	23	57	43	57
22	31	72	50	50	52	55	49	22	78
23	14	84	72	28	53	28	52	37	62
24	19	83	67	33	54	48	46	24	76
25	62	57	21	79	55	42	46	26	74
26	47	54	28	72	56	28	43	31	69
27	21	76	60	40	57	33	35	24	[illegible]
28	70	58	17	83	58	14	29	25	75
29	27	72	52	48	59	17	34	28	72
30	58	59	25	75	60	17	27	22	78

* $PDI_2 = 1 - d_{NP}$

** $E_2 = 1 - d_{NP}/d_P$

Table B4

Test 4 - Difficulty Coefficients Under P and NP Conditions (d_P, d_{NP}), Passage Dependency Index 2* and Passage Dependency Efficiency Index (E_2)**

Item	E_2	d_P	d_{NP}	PDI_2	Item	E_2	d_P	d_{NP}	PDI_2
1	47	98	51	49	23	70	60	18	82
2	49	76	39	61	24	08	78	72	28
3	15	92	79	22	25	12	80	70	30
4	14	94	81	19	26	36	72	46	54
5	40	80	48	52	27	12	52	46	55
6	33	92	61	39	28	17	55	46	54
7	37	75	47	53	29	72	55	16	84
8	08	88	81	19	30	17	69	57	43
9	80	84	17	83	31	38	39	25	76
10	08	68	74	26	32	12	74	83	17
11	38	40	25	75	33	14	56	63	37
12	16	84	71	29	34	35	64	41	59
13	20	46	37	63	35	77	60	14	86
14	37	84	53	47	36	14	44	50	50
15	06	47	50	51	37	36	83	45	55
16	14	83	72	29	38	03	66	68	32
17	07	53	49	51	39	71	62	18	82
18	13	76	66	34	40	07	56	60	41
19	70	71	21	79	41	12	49	43	57
20	18	47	38	62	42	06	61	57	43
21	47	77	41	59	43	12	67	75	25
22	35	60	39	61	44	21	48	38	62
					45	06	40	38	62

* $PDI_2 = 1-d_{NP}$

** $E_2 = 1-d_{NP}/d_P$

Table B5

Test 5 - Difficulty Coefficients Under P and NP Conditions (d_P, d_{NP}), Passage Dependency Index 2* and Passage Dependency Efficiency Index (E_2)**

Item	E_2	d_P	d_{NP}	PDI_2	Item	E_2	d_P	d_{NP}	PDI_2
1	07	86	80	20	24	23	69	53	47
2	14	88	76	24	25	28	45	33	68
3	46	55	81	19	26	51	65	32	69
4	04	49	51	49	27	02	51	50	50
5	43	71	41	60	28	08	85	78	22
6	52	84	41	59	29	12	69	60	40
7	21	88	69	31	30	41	45	26	74
8	18	83	69	32	31	56	44	19	81
9	07	70	65	35	32	19	59	48	52
10	22	79	61	39	33	49	69	35	65
11	25	79	60	40	34	78	49	11	89
12	42	40	24	77	35	13	30	26	74
13	56	65	29	71	36	44	40	22	78
14	09	90	83	18	37	68	69	22	78
15	28	85	62	38	38	35	64	42	58
16	08	86	79	21	39	49	67	34	66
17	81	59	11	89	40	10	41	37	63
18	64	53	19	81	41	47	41	22	78
19	38	85	53	47	42	48	44	23	77
20	46	74	40	60	43	20	21	25	[illegible]
21	13	81	70	30	44	22	71	55	[illegible]
22	14	61	53	47	45	60	59	24	76
23	50	76	38	63					

* $PDI_2 = 1 - d_{NP}$

** $E_2 = 1 - d_{NP}/d_P$

Table B6

Test 6 - Difficulty Coefficients Under P and NP Conditions (d_P, d_{NP}), Passage Dependency Index 2* and Passage Dependency Efficiency Index (E_2)**

Item	E_2	d_P	d_{NP}	PDI_2	Item	E_2	d_P	d_{NP}	PDI_2
1	92	89	07	93	22	09	80	73	27
2	65	89	31	69	23	45	78	42	58
3	06	78	82	18	24	02	54	53	47
4	04	86	90	10	25	68	77	24	76
5	96	65	03	98	26	27	78	57	43
6	07	88	83	17	27	11	61	54	46
7	11	75	67	33	28	43	61	35	66
8	44	86	48	52	29	39	36	22	78
9	08	89	81	19	30	11	81	73	28
10	08	51	47	53	31	40	34	21	79
11	14	50	43	57	32	20	62	50	50
12	09	70	63	37	33	52	73	35	65
13	10	73	65	35	34	67	55	18	82
14	19	39	32	69	35	42	40	56	44
15	56	78	34	66	36	17	46	38	62
16	15	63	54	47	37	54	32	15	86
17	24	69	52	48	38	22	55	43	57
18	42	75	44	56	39	05	37	35	65
19	46	79	43	57	40	16	47	39	61
20	64	50	18	82	41	16	55	46	54
21	07	91	85	15	42	03	32	32	68

* $PDI_2 = 1-d_{NP}$

** $E_2 = 1-d_{NP}/d_P$

APPENDIX C

Statistics for Passage Dependency of Test Items

Appendix C

Statistics for Passage Dependency of Test Items

Introduction

Tests of reading comprehension purport to measure how well a student understands what he is reading. Many of these tests employ questions to ascertain the degree of this understanding. At face value, these tests are very similar to any achievement test using the familiar multiple-choice format. In the case of reading comprehension tests, however, the tacit assumption exists that there is a direct relationship between the reading of the passage or the story and the ability to answer questions about it. In the case of a great many reading test items from standardized tests, this is a faulty assumption. It has been well demonstrated that the probability of a correct answer prior to reading the paragraph exceeds chance in the case of most reading comprehension questions. (Preston, 1964; Farr and Smith, 1970; Bickley, Weaver and Ford, 1968; Mitchell, 1967; Tuinman, 1970 ; Weaver and Bickley, 1967).

It must be pointed out that items with a relatively high passage independency (i.e., answerability with no passage being read) are not necessarily invalid. A student faced with answering such an item may actually use the information in the passage (which would be available under normal test conditions), even if he could have answered it by relying on extrinsic information such as general knowledge, syntactic cues, information present due to particular item sequences and the like.

The extent to which students will skip passages in actual testing situations is largely an unknown factor. Indirect evidence supporting the assumption that students can be tempted not to utilize information present in the passage is provided by Tuinman, 1972a and Tuinman (1972b).

The presence of passage independency in a reading comprehension test thus creates uncertainty about the validity of any measurement taken with this test. The problem is complicated by the fact that it is not at all clear to what extent the ability to answer questions without having read a passage is related to the ability to answer questions after reading a passage. (Preston, 1964; Tuinman, 1970 and Eurich, 1931).

The thesis of the present paper is that a "good" reading comprehension item need not only meet such generally accepted criteria as adequate item difficulty and item reliability, but that such items, in addition, must be tested against criteria derived from the necessity to maximize passage dependency in order to reduce uncertainty about the content validity of a specific measurement. Currently, test developers tend not to apply such criteria and little discussion of them is available in the literature. The remainder of this paper is devoted 1) to a description of a procedure to estimate the degree of "validity-uncertainty" of a test (or an item); 2) to illustrations using fictitious and actual item data and 3) to a description of the problems and assumptions associated with this procedure.

Some basic assumptions and formulas

It might be argued that all one needs to know about the passage dependency of a reading comprehension item is the probability of answering it correctly when no passage is presented. If this probability does not exceed chance (usually $1/k$, where k = number of options in a multiple-choice item), so goes the argument, the item is "good" in respect to passage dependency. This approach is too simplistic, however, as is quickly demonstrated by the item which has a difficulty of $1/k$ in both the passage present and the passage absent conditions. The following discussion will point out further complications. First, a number of statistics should be defined:

p_a = the proportion of correct responses to item i under the no-passage (NP) condition

$p_b = 1 - p_a$ the proportion of incorrect responses to item i under the NP condition

p_c = the proportion of correct responses to item i under the passage (P) condition

$p_d = 1 - p_c$, the proportion of incorrect responses to item i under the P condition

To obtain p_a and p_b, the items are administered to a sample from a given population of test takers who answer the questions without being able to read the passages on which the questions are based. The statistics p_c and p_d are obtained from an independent sample of subjects from the same population. Marks and Noll (1967) obtained estimates of the contribution of passages to items based on them by administering the items twice to the same group of respondents. The procedure proposed here avoids both the measurement and practical problems associated with this approach.

Ss who answer the items under the NP condition must either act on the basis of some extrinsic information (i.e., information not derived from the passage) or they are guessing. Accordingly, we may write:

$$p_a = p_{a1} + p_{a2} \text{ where,} \qquad 1)$$

p_{a1} = the proportion of correct responses to item i based on guessing

p_{a2} = the proportion of correct responses based on extrinsic information.

Due to the peculiarities of the NP-condition some behaviors which are normally considered to be part of the "guessing" behavior (Tinkelman, 1971) are currently included under the category of responding on the basis of extrinsic information--for example, making use of semantic or syntactic cues available from the wording of the question. It is necessary to assume that there is merit in the use of conventional correction formulas for guessing. The likelihood of this assumption being correct is supported by the fact that many essentially non-guessing behaviors are excluded from the guessing component under the definitions employed here. Later in the paper this issue is discussed further; for the moment it is assumed that p_{a1} can be estimated from the proportion of wrong answers (p_b or $1 - p_a$), using the logic of a correction for guessing formula. Thus,

$$p_{a1} = p_b/(k-1) = (1-p_a)/(k-1) \qquad 2)$$

where k = the number of options per item.

It then follows that p_{a2} or the proportion of correct responses based on extrinsic information is given by 3)

$$p_{a2} = p_a - p_{a1} \quad 3)$$

Expressed directly in terms of p_a, p_{a2} is given by 4)

$$p_{a2} = k(p_a) - 1 /(k-1) \quad 4)$$

Analogous to the partitioning followed above, the statistics obtained from the administration of the test items under the P-condition result in:

p_c = the proportion of correct responses

p_{c1}= the proportion of correct responses based on guessing

p_{c2}= the proportion of correct responses based on information extrinsic to the passage

p_{c3}= the proportion of correct responses based on the passage.

Thus,

$$p_c = p_{c1} + p_{c2} + p_{c3} \quad 5)$$

Again, p_{c1} can be calculated from the proportion of incorrect responses, $p_d = 1 - p_c$. Thus,

$$p_{c1} = (1-p_c)/(k-1) \quad 6)$$

Since the passage usually does provide at least some information not contained in the questions only, $p_{a1} \neq p_{c1}$, except under the rare condition that $p_a = p_c$.

It is impossible to partition p_{c2} and p_{c3}, given only proportional data p_a, p_b, p_c, and p_d. This means that no direct estimate of the contribution of the passage to the responses to an item or set of items is available. However, it is possible to calculate in a straightforward manner a number of statistics which may be of use in deciding on the quality of the item or group of items under consideration.

These statistics are:

a) E_{max}, the maximum contribution of extrinsic information under the P-condition

b) P_{min}, the minimum contribution of the passage

c) P_{max}, the maximum contribution of the passage

If none of the students responding to an item is able to utilize any of the information in the passage, extrinsic information will be exercising maximum influence. This maximum influence is indexed by p_{a2}. Thus,

$$E_{max} = p_{a2} \qquad 7)$$

The minimum contribution of the passage is given by 8)

$$P_{min} = p_c - p_{c1} - p_{a2} \qquad 8)$$

Equation 8) follows from 5) given the identity $p_{a2} = p_{c2}$, which holds in the case of minimum contribution of the passage.

P_{min} can be directly calculated from p_a and p_c using 9)

$$P_{min} = k(p_c - p_a)/(k-1) \qquad 9)$$

The maximum contribution of the passage occurs only when the students do not utilize any extrinsic information when answering the items, hence when $p_{c2} = 0$. In this case the value of p_{c3} equals

$$P_{max} = p_c - p_{c1}, \text{ or} \qquad 10)$$

$$P_{max} = (k(p_c) - 1)/(k-1) \qquad 11)$$

It may be redundant to observe that

$$E_{max} = P_{max} - P_{min} \qquad 12)$$

Since the statistics calculated above are all linear transformations of the basic probabilities p_a and p_c, upper and lower confidence limits for these probabilities can be substituted in the formulas above in order to arrive at confidence regions for the various statistics of interest. Furthermore, a correction for the calculation of p_{a1} and p_{c1} must be applied in the case of omitted responses.

An illustration with fictitious data.

Insert Table C7 about here

In Table C7, hypothetical values of p_a and p_c are paired and the values of the corresponding E_{max}, P_{min} and P_{max} are presented. First, from equation 4) it is clear that E_{max} depends solely on the value of p_a. In a sense, therefore, this statistic does not contribute any new knowledge about an item. E_{max} is useful as an indication of the maximum proportion of correct responses to an item under the P-condition which could be attributed to extrinsic knowledge. If, for instance, under the NP-condition 40% of the Ss answer item 1 correctly, this does not mean that under the P-condition 40% of the correct responses could be due to extrinsic information. Rather, this proportion cannot be higher than .20 as can be seen in Table C7. The P_{max} statistic fulfills a similar function. It is exclusively dependent on p_c and therefore adds no information in terms of relative relations among the items. Yet, P_{max} is useful as an index of the absolute proportion of correct responses under influence of the passage.

It was noted above that a determination of an item's quality in terms of passage dependency can not be made only on the basis of p_a. Given a multiple-choice test, a number of items may have a p_a value of .25. The p_a statistic is simply of no help in discriminating among these items in terms of the extent to which the passage controls the responses. In this particular case, either P_{min} or P_{max} provides the information necessary to make a decision about an item's desirability. For norm-referenced measurement an item with p_a = .25 and p_c = .60 might be a desirable item, whereas for criterion-referenced measurement the combination p_a = .25 and p_c = 1.00 might be preferable.

Table C7 shows negative values of P_{min}. It seems illogical to talk about negative minimal contributions of the passage in answering questions based on that passage. Yet, these negative P_{min}'s are realistic and index a situation which is not uncommon. Some reading comprehension items are easier with no passage present than with the passage present. This is true, of course, for miskeyed items, but is also true when a passage contains ambiguous information or when it leads some students not to choose the obviously right answer. This case is illustrated by the second set of p_a and p_c values in Table C7.

A p_{min} equal to -.20 indicates that at the most 20% of the incorrect responses would be a function of misleading information in the passage. This interpretation of negative P_{min}'s is illustrated best by the item with p_a = 1.00 and p_c = .25. Here 25% of the responses would be correct presumably by guessing and all of the remaining responses are wrong as a function of ambiguous or misleading information in the passage.

It may be further observed that P_{min} can only equal 1.00 if p_a = .25 and p_c = 1.00. In this case, the minimum and maximum contribution of the passage are identical. However, P_{max} can be 1.00 while P_{min} = 0.

This is the case when both p_a and p_c are equal to 1.00. These observations indicate an important point. It was argued earlier that a p_a value higher than $1/k$ (where k = number of options) does not necessarily mean that an item is invalid, i.e., that responding to it under the P-condition does not involve utilization of the information from the passage. What matters is that uncertainty about how the item was answered, (i.e., whether passage information was used or not) should be minimal. This uncertainty is indexed directly by the difference between P_{min} and P_{max}, and therefore by E_{max}. The smaller E_{max}, the less uncertainty about the actual degree to which the passage influenced the responses. Yet, as the second set of five items in Table C7 indicate, minimal E_{max}'s are possible under a variety of conditions; zero uncertainty is possible both with a 0.0 and 1.00 minimal contribution of the passage. Relatively low E_{max}' are only meaningful in relation to their corresponding P_{min} values. This particular point will be further illustrated in the next section of this paper.

<u>An illustration with real data</u>

The statistics described above were calculated for six different standardized tests of reading comprehension, all of which are frequently used in routine assessment of reading performance in the public schools. Four of the tests were administered to a sample of 4th, 5th and 6th graders. One test was suited for the 4th grade exclusively and one test was administered in only the 5th and 6th grades. In the NP-condition each grade provided 400 subjects. In the P-condition 200 subjects per grade and per test were used. The total sample of subjects used was slightly over 9,000. A full description of the sample and the procedures followed is provided in the main body of this report. Table C8 contains the validity statistics calculated for these tests. The entries are averages for the number of items indicated. In some cases this number is smaller than those present in the test. This is a function of the fact that only the items completed by all the Ss in the P-condition were included in the present calculations.

Insert Table C8 about here

A number of observations need to be made. Imagine that the decision called for is to select the test that gives the most guarantees that the responses to the items are a function of information in the passage, at the same time not violating other rationales for selecting tests. First, the test difficulties do not vary much with the exception of Test 1. This test is the Nelson Reading test, rather a speed test, and since only the first 40 items are considered here, the low difficulty (.77) is not surprising. None of the tests fares very well on the question of passage dependency of the items. The best in the group, Test 2 (California Achievement Battery Reading, Level 3A) has an average difficulty under the NP-condition of .34. This means that, at most, 14% of the correct responses under the P-condition could be due to extrinsic information. Test 3 (a subset of SRA Reading Test items) produces about as much guaranteed responding under influence of the passage as does Test 2 (.37 versus .38) but allows for the possibility that the passage may determine a higher percentage of correct responses (.59 to .52). The price for buying more passage control by selecting Test 3 as the first choice, however, is more uncertainty about the validity of the responses of a given set of respondents. The comparison of Tests 1 and 6 illustrates another selection problem. Both have identical E_{max} values. However, Test 1 seems preferable in view of its higher P_{min} value, provided that the low difficulty level of the test does not constitute an a priori reason to reject the test. This, it appears, would depend on other criteria set for test usage, criteria unrelated to passage control over the responses.

The necessity to relate E_{max} values to P_{min} values as illustrated above suggests the use of a P_{min}/E_{max} ratio.

Insert Table C9 about here

Table C9 illustrates the use of the validity indices proposed here for the selection of items, as opposed to those for tests. Included also is the ratio of P_{min} to E_{max}. Briefly, Table C9 suggests the following observations:

1. The difference between p_c and p_a is not enough to determine the quality of an item in terms of passage dependency characteristics. Compare Items 1 and 41, for example, on Test 1.

2. Negative E_{max} values arise when relatively easy items contain a false option which seems a very good choice when there is no information from the passage present. This leads to artifically inflated P_{min} values. The actual contribution of the passage in these cases is given by P_{max}, with zero uncertainty about the passage's contribution existing. (See Item 18, Test 1.)

3. Items of equal difficulty under the P-condition may differ vastly in difficulty under the NP-condition. (Compare Items 8 and 9, Test 2). These differences are most meaningfully reflected in the P_{min} values of the items.

4. Nearly identical E_{max} values may obscure differences between items reflected in the P_{min}/E_{max} ratio (Item 1, Test 1 and Item 8, Test 2). Given the acceptability cf a p_c value equal to .95, Item 1 is to be preferred over Item 8 on grounds of the higher ratio.

Table C7 indicated that for a given p_a-value, P_{min} and P_{max} are highly correlated, whereas the correlation between E_{max} and these two statistics is zero (in absence of variance among E_{max} values for a particular p_a-value). In a set of items with varying combinations of p_a and p_c values, however, the relations are less simple. Table C10 contains a correlation matrix based on the completed items for Test 5. This matrix is representative of those computed for the other tests.

Insert Table 10 about here

It must be borne in mind that since E_{max} is a linear transformation of p_a, and P_{max} is such a transformation of p_c, correlations between E_{max} and any variable and between P_{max} and any variable also hold for that variable and p_a and p_c, respectively. The most important conclusion to be drawn from Table C10 is that, for a particular set of items, each of the variables calculated provides different information relating to the issue of passage dependency of a particular item.

Discussion

The major point brought out by the preceding analysis is that it is insufficient to consider the passage dependency of reading comprehension items solely in terms of their difficulty under the NP-condition. No attempt has been made to provide a step-by-step procedure for evaluation of items in this respect. Rather, a number of indices have been discussed which under certain conditions may be of use in comparing items or tests in terms of their measuring behaviors which are under control of the information in the passage.

The fact that the statistics proposed are linear transformations of p_a, p_c and p_c-p_a indicates that for practical decisions the latter quantities might be used just as well P_{min}, P_{max} or E_{max}. For example, in many instances it might suffice to look for items with p_c values within an acceptable range on large values of p_c-p_a.[1] The reason for considering the three new statistics proposed here is, first of all, the fact that they emphasize an essential problem with reading comprehension items (i.e., the contribution of the passage to the probability of a correct response). Secondly, the statistics allow an expression of the limits of this contribution in terms of percentages which are considered to be more meaningful by this author than the mere proportions of correct responses under the P and NP-conditions.

It already has been acknowledged that the adequacy of the analysis outlined above depends on the degree to which application of the correc-

---

[1] I wish to thank Dr. Robert Linn, ETS, for his suggestions on this matter

tion-for-guessing formula may be assumed to be correct. This assumption bears directly on the calculation of p_{al} and p_{cl}. As Tinkelman (1971) points out, the term "guessing" quite loosely refers to an array of behaviors, many of which involve responding in terms of partial information rather than essentially non-systematic response behaviors. To the degree that this is the case, the application of correction-for-guessing formula's is questionable. In terms of the present analysis this conclusion must be interpreted in the light of the following observations.

First, the concept of "extrinsic information" includes all utilization of "partial" information which normally would be thought of as part of "guessing." The residual behaviors included in "guessing" are more likely to be of a non-systematic nature than is the case in those applications where no attempt is made to separate the various components of behaviors leading to incorrect responses.

Secondly, the formulas and concepts developed above may be used with entire tests as well as with individual items. It stands to reason that the various estimates of responses based on a particular source of information are more reliable in the former case. This is true, too, in the case of the statistics yielded by the application of the correction-for-guessing formula. The assumption underlying this application in the case of a set of items is a weaker assumption. It is only necessary to assume that the sums of the responses for the wrong options are equal across items and subjects. That is an assumption which is more likely to hold than its more stringent counterpart which must hold in the case of the application of the formula for correction with individual items: an equal number of respondents selected each of the k-1 wrong options.

Thirdly, the analysis of reading comprehension items presently is in a rather primitive stage. If a more adequate analysis calls for distractors that are basically equipotent, a very basic tenet of sound item writing is merely reiterated. Currently, reading comprehension items too often contain distractors with very low potency. Such distractors not only lower the general efficiency of these items, but, as shown, actively interfere with adequate assessment of these item's, passage dependency characteristics since lack of equipotency of distractors reduces non-systematic selection of options.

References

Bickley, A. C., Weaver, W. W., Ford, F. Information removed from multiple-choice item responses by selected grammatical categories. Psychological Reports, 1968, 23, 613-14.

Eurich, A. C. A method for measuring retention in reading. Journal of Educational Research, 1931, 24, 202-8.

Farr, R. and Smith, C. B. The effects of test item validity on total test reliability and validity. In G. Schick and M. M. May (Eds.). Reading: Process and Pedagogy. Nineteenth Yearbook of the National Reading Conference, Vol. 1, Milwaukee, Wisconsin, 1970, 122-134.

Marks, E. and Noll, G. Procedures for evaluating reading and listening comprehension tests. Educational and Psychological Measurement, 1967, 335-343.

Mitchell, R. W. A comparison of children's responses to an original and experimental form of subtests GS and NP of the Gates Basic Reading Tests. Unpublished Doctoral Dissertation, University of Minnesota, 1967.

Preston, R. C. Ability of students to identify correct responses before reading. Journal of Educational Research, 1964, 58, 181-183.

Tinkelman, S. N. Planning the objective test. In Thorndike, R. L. (Ed.) Educational Measurement, 2nd Edition. Washington, D.C.: American Council on Education, 1971, 46-80.

Tuinman, J. J. Selected aspects of the assessment of the acquisition of information from reading passages. Unpublished Doctoral Dissertation, University of Georgia, 1970.

Tuinman, J. J. Inspection of reading comprehension passages as a function of passage dependency of test items. Research Report of the Institute for Child Study, Indiana University, Bloomington, Indiana, 1972a.

Tuinman, J. J. Children's willingness to skip reading passages when taking reading comprehension tests. Southern Journal of Educational Research, 1972b, 6, 1-13.

Weaver, W. W. and Bickley, A. C. Sources of information for responses to reading test items. Proceedings of the 75th Annual Convention of the American Psychological Association, 1967, 2, 293-294.

Table C1

Test 1 - Values of Item Validity Statistics Described in Appendix C for all Items.

Item	P_{max}	P_{min}	P_{min}/E_{max}	E_{max}
1	.93	.17	.23	.76
2	.85	.77	10.56	.07
3	.90	.56	1.63	.34
4	.92	.12	.15	.80
5	.88	.61	2.31	.26
6	.83	.11	.16	.72
7	.86	.63	2.74	.23
8	.85	.24	.40	.61
9	.74	.45	1.51	.30
10	.78	.89	-8.10	-.11
11	.80	.28	.54	.52
12	.84	.32	.62	.52
13	.78	.21	.38	.56
14	.87	.12	.16	.74
15	.87	.45	1.10	.41
16	.84	.32	.63	.52
17	.83	.37	.79	.46
18	.70	.93	-4.05	-.23
19	.59	.35	1.44	.24
20	-.24	-.33	-3.57	.09
21	.84	.46	1.25	.37
22	.83	.69	5.19	.13
23	.59	.84	-3.37	-.25
24	.57	.18	.45	.40
25	.67	.83	-5.20	-.16
26	.51	.49	24.73	.02
27	.68	.62	10.40	.06
28	.68	.31	.84	.37
29	.65	.71	-11.76	-.06
30	.49	.26	1.13	.23
31	.81	.39	.92	.42
32	.68	.41	1.54	.27
33	.76	.33	.79	.43
34	.41	.22	1.10	.20
35	.61	.28	.87	.33
36	.45	.49	-15.17	-.03
37	.47	.41	6.98	.06
38	.67	.38	1.29	.29
39	.67	.47	2.32	.20
40	.40	.48	-5.58	-.09
41	.23	.15	1.84	.08
42	.43	.35	4.34	.08
43	.17	.15	8.85	.02
44	.39	.53	-3.94	-.13
45	.27	.41	-2.94	-.14
46	.36	.22	1.59	.14
47	.40	.39	36.38	.01
48	.58	.45	3.47	.13
49	.48	.24	.97	.25
50	.42	.46	-10.61	-.04
51	.54	.20	.58	.34
52	.49	.24	.94	.25
53	.54	.28	1.08	.26
54	.33	.32	121.00	.00
55	.38	.26	2.24	.12
56	.48	-.07	-.12	.55
57	.46	-.06	-.11	.52
58	.11	.20	-2.2[illegible]	-.09
59	.17	.15	8.6[illegible]	.02
60	.20	.18	9.50	.02
61	.10	.17	-2.56	-.07
62	.30	-.01	-.05	.32
63	.19	-.02	-.11	.21
64	.16	.18	-10.38	-.02
65	.16	.26	-2.57	-.10
66	-.13	-.04	.36	-.10
67	-.02	.01	-.36	-.03
68	.18	.01	.06	.17
69	.22	-.10	-.32	.32
70	-.17	-.01	.07	-.16
71	.21	.31	-2.88	-.11
72	.09	-.20	-.69	.29
73	-.10	-.15	-2.87	.05
74	.004	.04	-.10	-.04
75	-.16	-.21	-3.72	.06

Table C2

Test 2 - Values of Item Validity Statistics Described in Appendix C for all Items.

Item	P_{max}	P_{min}	P_{min}/E_{max}	E_{max}	Item	P_{max}	P_{min}	P_{min}/E_{max}	E_{max}
1	.92	1.07	-6.98	-.15	22	.02	.01	1.60	.01
2	.91	1.11	-5.41	-.21	23	-.09	-.15	-2.58	.06
3	.66	.48	2.81	.17	24	.78	.19	.32	.59
4	.75	.66	7.73	.09	25	.68	.16	.32	.52
5	.49	.54	-9.93	-.05	26	.80	.88	-11.36	-.08
6	.62	.70	-8.12	-.09	27	.71	.08	.13	.63
7	.76	.65	5.44	.12	28	.62	.43	2.31	.19
8	.74	.52	2.34	.22	29	.43	.06	.17	.37
9	.80	.92	-7.64	-.12	30	.31	-.07	-.19	.39
10	.77	.77	.50	-.00	31	.15	.14	8.67	.02
11	.59	.41	2.38	.17	32	.38	.32	5.24	.06
12	.57	.38	2.01	.19	33	.23	.06	.33	.18
13	.24	.10	.78	.13	34	.37	.05	.17	.32
14	.79	.63	3.92	.16	35	.68	.57	5.22	.11
15	.46	.21	.88	.25	36	.70	.65	15.38	.04
16	.65	.39	1.56	.25	37	.55	.52	13.73	.04
17	.72	.40	1.25	.32	38	.56	.62	-10.57	-.06
18	.61	.32	1.09	.29	39	.44	.47	-14.50	-.03
19	.45	.17	.60	.28	40	.39	.37	18.69	.02
20	.25	.18	2.32	.08	41	.12	.10	4.39	.02
21	.17	-.02	-.10	.19	42	.11	.11	16.80	.01

Table C3

Test 3 - Values of Item Validity Statistics Described in Appendix C for all Items.

Item	P_{max}	P_{min}	P_{min}/E_{max}	E_{max}	Item	P_{max}	P_{min}	P_{min}/E_{max}	E_{max}
1	.93	.45	.92	.49	31	.48	.60	-5.21	-.11
2	.89	.14	.18	.76	32	.42	.37	7.08	.05
3	.78	.33	.72	.45	33	.39	.33	5.33	.06
4	.80	.28	.54	.52	34	[illegible]	.27	.62	.44
5	.84	.99	-6.85	-.14	35	.59	.67	-8.40	-.08
6	.67	.32	.92	.35	36	.61	.52	6.19	.09
7	.68	.29	.73	.39	37	.23	.23	-19.7[illegible]	-.01
8	.72	.47	1.89	.25	38	.46	.45	114.33	.01
9	.25	.28	-9.17	-.03	39	.20	.09	.93	.10
10	.89	.59	1.91	.31	40	.43	.29	2.17	.13
11	.52	.54	-22.61	-.02	41	.67	.71	-19.74	-.04
12	.51	.30	1.46	.20	42	.24	.06	.37	.17
13	.52	.49	15.83	.03	43	.46	.39	5.88	.07
14	.84	.17	.25	.67	44	.28	.24	5.53	.04
15	.57	.51	8.15	.06	45	.24	.28	-7.31	-.04
16	.55	.22	.66	.33	46	.52	.07	.16	.45
17	.46	.01	.01	.45	47	.39	.36	10.27	.03
18	.43	.35	3.87	.09	48	.17	.23	-3.93	-.03
19	.38	.32	5.71	.06	49	.54	.54	-204.00	-.00
20	.76	.73	23.78	.03	50	.37	.30	4.37	.07
21	.74	.47	1.69	.28	51	.42	.18	.72	.24
22	.63	.30	.91	.33	52	.32	.36	-9.07	-.04
23	.79	.16	.26	.63	53	.37	.19	1.14	.17
24	.77	.21	.37	.56	54	.28	.30	-20.36	-.01
25	.42	.47	-9.57	-.05	55	.27	.26	14.77	.02
26	.38	.34	7.70	.04	56	.24	.17	2.10	.03
27	.68	.21	.45	.47	57	.14	.16	-10.72	-.01
28	.44	.55	-5.24	-.10	58	.05	.05	-41.00	-.00
29	.62	.26	.72	.36	59	.12	.08	1.97	.04
30	.45	.46	-85.50	-.01	60	.03	.06	-1.73	-.03

Table C4

Test 4 - Values of Item Validity Statistics Described in Appendix C for all Items.

Item	P_{max}	P_{min}	P_{min}/E_{max}	E_{max}	Item	P_{max}	P_{min}	P_{min}/E_{max}	E_{max}
1	.97	.62	1.75	.35	24	.71	.08	.13	.63
2	.68	.49	2.65	.19	25	.73	.13	.22	.60
3	.90	.19	.26	.71	26	.63	.35	1.26	.2[illegible]
4	.92	.17	.23	.75	27	.36	.08	.30	.27
5	.73	.42	1.36	.31	28	.40	.12	.44	.28
6	.89	.40	.83	.49	29	.40	.53	-4.19	-.13
7	.67	.37	1.28	.29	30	.58	.15	.35	.43
8	.84	.09	.12	.75	31	.19	.20	-29.60	-.01
9	.79	.90	-8.43	-.11	32	.66	-.11	-.15	.77
10	.58	-.08	-.12	.65	33	.41	-.10	-.20	.51
11	.20	.21	-77.50	-.00	34	.51	.30	1.4[illegible]	.21
12	.79	.18	.29	.61	35	.47	.62	-4.12	-.15
13	.28	.13	.81	.15	36	.25	-.08	-.24	.34
14	.79	.42	1.14	.37	37	.11	-.16	-.59	.27
15	.29	-.04	-.12	.33	38	.55	-.03	-.05	.58
16	.78	.16	.26	.62	39	.49	.58	-6.15	-.09
17	.37	.05	.16	.32	40	.40	-.05	-.11	.46
18	.68	.13	.24	.55	41	.32	.08	.32	.25
19	.61	.66	-12.40	-.05	42	.48	.05	.11	.43
20	.29	.11	.62	.18	43	.56	-.11	-.16	.67
21	.70	.48	2.21	.22	44	.31	.13	.76	.17
22	.46	.28	1.50	.19	45	.20	.03	.20	.17
23	.46	.56	-5.75	-.10					

Table C5

Test 5 - Values of Item Validity Statistics Described in Appendix C for all Items.

Item	P_{max}	P_{min}	P_{min}/E_{max}	E_{max}	Item	P_{max}	P_{min}	P_{min}/E_{max}	E_{max}
1	.81	.07	.10	.74	24	.58	.21	.56	.37
2	.84	.17	.25	.67	25	.27	.17	1.71	.10
3	.40	-.34	-.46	.74	26	.53	.45	5.06	.09
4	.32	-.03	-.08	.35	27	.34	.01	.04	.3[illegible]
5	.60	.40	1.95	.21	28	.80	.09	.13	[illegible]
6	.79	.58	2.78	.21	29	.58	.11	.23	.47
7	.83	.24	.41	.59	30	.27	.25	13.36	.02
8	.78	.20	.34	.58	31	.25	.33	-4.32	-.08
9	.60	.07	.13	.53	32	.45	.15	.48	.30
10	.72	.23	.48	.48	33	.58	.45	3.47	.13
11	.72	.26	.56	.46	34	.33	.51	-2.73	-.19
12	.20	.22	-11.20	-.02	35	.06	.05	5.71	.01
13	.53	.48	9.47	.05	36	.19	.23	-5.87	-.04
14	.87	.10	.13	.77	37	.59	.62	-17.96	-.03
15	.81	.31	.64	.49	38	.52	.29	1.31	.23
16	.82	.01	.13	.72	39	.56	.43	3.58	.12
17	.45	.64	-3.46	-.18	40	.21	.05	.33	.16
18	.38	.45	-5.78	-.08	41	.21	.25	-5.94	-.04
19	.80	.43	1.15	.37	42	.25	.28	-8.79	-.03
20	.65	.46	2.30	.20	43	-.05	-.06	-14.00	.004
21	.74	.14	.24	.60	44	.62	.21	.52	.41
22	.48	.11	.31	.37	45	.45	.47	-25.29	-.02
23	.67	.51	3.04	.17					

Table C6

Test 6 - Values of Item Validity Statistics Described in Appendix C for all Items.

Item	P_{max}	P_{min}	P_{min}/E_{max}	E_{max}	Item	P_{max}	P_{min}	P_{min}/E_{max}	E_{max}
1	.85	1.03	-4.54	.24	22	.73	.10	.16	.64
2	.85	.77	9.97	.08	23	.70	.47	2.04	.23
3	.70	-.06	-.08	.76	24	.39	.01	.04	.37
4	.81	-.05	-.06	.86	25	.69	.70	-75.29	-.01
5	.53	.83	-2.78	-.30	26	.71	.29	.68	.42
6	.85	.08	.10	.77	27	.48	.09	.24	.39
7	.67	.11	.19	.56	28	.48	.35	2.76	.13
8	.81	.51	1.66	.31	29	.15	.19	-4.90	-.04
9	.85	.10	.14	.74	30	.75	.12	.18	.63
10	.35	.05	.18	.29	31	.13	.18	-3.14	-.06
11	.33	.10	.40	.24	32	.49	.16	.49	.33
12	.59	.09	.17	.51	33	.64	.51	3.88	.13
13	.63	.10	.18	.53	34	.40	.49	-5.36	-.09
14	.18	.10	1.11	.09	35	.19	-.22	-.54	.41
15	.70	.58	4.91	.12	36	.28	.11	.61	.18
16	.51	.13	.34	.38	37	.09	.23	-1.62	-.14
17	.59	.22	.61	.37	38	.41	.17	.69	.24
18	.67	.42	1.65	.25	39	.15	.03	.18	.13
19	.72	.48	2.05	.23	40	.29	.10	.53	.19
20	.34	.43	-4.76	-.09	41	.39	.12	.43	.28
21	.88	.08	.10	.30	42	.09	-.01	-.11	.10

Table C7

Illustrative Validity Statistics for Selected Values of p_a and p_c

p_a Value	p_c	E_{max}	P_{min}	P_{max}
p_a = .25	p_c = .25	.00	.00	[illegible]
	p_c = .40	.00	.20	.20
	p_c = .60	.00	.47	.47
	p_c = .80	.00	.73	.73
	p_c = 1.00	.00	1.00	1.00
p_a = .40	p_c = .25	.20	-.20	.00
	p_c = .40	.20	.00	.20
	p_c = .60	.20	.27	.47
	p_c = .80	.20	.53	.73
	p_c = 1.00	.20	.80	1.00
p_a = .50	p_c = .25	.33	-.33	.00
	p_c = .40	.33	-.13	.20
	p_c = .60	.33	.14	.47
	p_c = .80	.33	.40	.73
	p_c = 1.00	.33	.67	1.00
p_a = 1.00	p_c = .25	1.00	-1.00	.00
	p_c = .40	1.00	- .80	.20
	p_c = .60	1.00	- .53	.47
	p_c = .80	1.00	- .27	.73
	p_c = 1.00	1.00	.00	1.00

Table C5

Mean Values of Validity Statistics for Six Tests

Test	Items	p_c	p_a	P_{min}	P_{max}	E_{max}
1	40	.77	.46	.42	.70	.28
2	42	.64	.34	.38	.52	.14
3	40	.69	.41	.37	.59	.21
4	45	.66	.50	.22	.54	.33
5	45	.64	.45	.25	.52	.27
6	40	.65	46	.26	.54	.28

Table C9

Validity Statistics for Selected Items

Item Number	p_c	p_a	P_{min}	P_{max}	E_{max}	P_{min}/E_{max}*
Test 1						
1	.95	.82	.17	.93	.76	.23
5	.91	.45	.61	.88	.26	2.32
18	.77	.07	.93	.70	-.23	-4.05
41	.43	.31	.15	.23	.08	1.84
Test 4						
8	.88	.81	.09	.84	.75	.12
9	.84	.17	.90	.79	-.11	- 8.43
10	.47	.50	-.04	.29	.33	- .12
31	.39	.25	.20	.19	-.01	-29.60

* This ratio was calculated with unrounded values of P_{min} and E_{max}.

Table C10

Correlation Matrix
Test 5 (N=45)

	(1) P_{min}	(2) E_{max}	(3) P_{max}
P_{min}	—		
E_{max}	-.56	—	
P_{max}	.18	.71	—
P_{min}/E_{max}	-.10	.29	.27