DOCUMENT RESUME

ED 222 549 TM 820 702

AUTHOR	Powers, Stephen
TITLE	The Effect of Testwiseness on the Reading Achievement Scores of Minority Populations. Final Report.
INSTITUTION	Tucson Unified School District, Ariz.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
PUB DATE	Mar 82
GRANT	NIE-G-80-0076
NOTE	98p.
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	American Indians; Black Students; *Elementary Education; Ethnic Groups; Hispanic Americans; Intentional Learning; *Racial Differences; *Reading Achievement; Reading Comprehension; Response Style (Tests); *Test Coaching; *Test Wiseness; White Students
IDENTIFIERS	California Achievement Tests; Comprehensive Tests of Basic Skills; Test of Testwiseness (Slakter)

ABSTRACT

Testwiseness (TW) and its effect on the reading comprehension scores in four ethnic populations (Black, Hispanic, Native American, and Anglo) from the Tucson Unified School District were investigated. Inservice-trained teachers presented seven classroom TW teaching and practicing sessions of 15 minutes each to students in grades 3, 5, and 7. Students were administered the Test of Testwiseness; they were pretested and posttested with the Reading Comprehension subtest of either the California Achievement Tests (grades 3 and 7) or the Comprehensive Tests of Basic Skills (grade 5). Differential reading ability and socioeconomic level, both variables highly correlated with testwiseness, were statistically controlled. It was found that (1) no ethnic group possessed a significantly greater or lesser amount of testwiseness when different reading ability and socioeconomic levels were controlled; (2) when instructed in testwiseness, gains appeared comparable among ethnic groups; and (3) the amount of testwiseness training did not appear to affect reading comprehension scores on a standardized achievement test. (Author/PN)

The Effect of Testwiseness on the Reading Achievement Scores of Minority Populations

Final Report to the
National Institute of Education
United States Department of Education

Stephen Powers, PhD., Principal Investigator
Department of Legal and Research Services
Tucson Unified School District
Tucson, Arizona 85719

NIE Teaching and Learning Program Grant
NIE-G-80-0076

March 1982

TABLE OF CONTENTS

	Page
Acknowledgement	i
Abstract	ii
Introduction	1
Method	3
Sample Selection	3
Instrumentation	5
Procedure	6
Results	8
Research Question 1	9
Research Question 2	13
Research Question 3	16
Research Question 4	20
Research Question 5	23
Research Question 6	23
Discussion	25
Appendix 1: Testwiseness Instruction	29
Appendix 2: The Development and Validation of a Group Test of Testwiseness	57
Appendix 3: Test of Testwiseness	69
Appendix 4: Preliminary Findings: Report to Participating Teachers and Principals	77
Appendix 5: Repeated Measures Analysis of Variance on Reading Comprehension Scores	93
References	97

ACKNOWLEDGEMENT

To the 854 students of the Tucson Unified School District who participated in this study, to the teachers who volunteered their time, to the principals who gave their permission for this study to be conducted, and to other administrators who supported this study - I would like to acknowledge my gratitude.

I would like to acknowledge the contribution to this project of Dr. Darrell L. Sabers, Professor of Educational Psychology, the University of Arizona, Tucson. His expertise in the areas of educational measurement, statistics and testwiseness provided inestimable depth to the present study. Moreover, his ready assistance far exceeded the monetary considerations of the proposal.

In addition, I would like to express my gratitude to Dr. Lewis Pike, National Institute of Education Project Officer, for his assistance and interest during the course of this study.

Stephen Powers

Stephen Powers, Principal Investigator

ABSTRACT

This study fills a need in the research literature on testwiseness because it focuses on (1) possible ethnic group differences in testwiseness, (2) possible differential gains of ethnic groups in testwiseness after instruction, and (3) possible effects of testwiseness on reading achievement test performance. This study included in its sample - Native American, Black, Hispanic and Anglo students in the third, fifth and seventh grades. Differential reading ability and socioeconomic level, both variables highly correlated with testwiseness, were statistically controlled.

Among the salient findings were the following: (1) No ethnic group possessed a significantly greater or lesser amount of testwiseness when different reading ability and socioeconomic levels were controlled, (2) When instructed in testwiseness, gains appeared comparable among ethnic groups, and (3) The amount of testwiseness training did not appear to affect reading comprehension scores on a standardized achievement test.

INTRODUCTION

The public's interest in testwiseness has been growing while articles in recent newspapers and journals depict this heightened interest. The Arizona Daily Star (1981) reported, "100 school districts join test-help plan". The Dallas Times Herald (Austin 1981) proclaimed, "Teachers pressured to teach test-taking skills". The American School Board Journal (Gifford and Fluitt, 1980) carried the article, "How to make your students test-wise", and the Journal of Reading carried an article by the Assistant Provost at the University of Maryland at College Park (McPhail, 1981), arguing for the teaching of test-taking skills.

Testwiseness (TW) was first identified as a possible effector of reliability by Thorndike (1951). A viable conceptual framework for this construct was provided by Millman, Bishop, and Ebel (1965) whose definition of TW is employed in this study:

> a subject's capacity to utilize the characteristics and formats of the test and/or the test taking situation to receive a high score.

Researchers have shown that TW can be measured (Gibb, 1964; Millman, 1966; Slakter et al., 1970a; Woodley, 1973; Bajtelsmit, 1975a), that TW can be taught (Gaines and Jongsma, 1974; Sarnacki, 1979), and that training in TW skills improves performance on tests designed to measure TW (Gibb, 1964; Slakter et al., 1970a; Moore, Schutz, and Baker 1966; Langer et al.,1973). More relevant to the purposes of this study, researchers have shown that instructions in TW raises the scores of students measured on standardized tests such as the Comprehensive Tests of Basic Skills (Gaines and Jongsma, 1974), the Stanford Reading Test (Callenbach, 1973), the Metropolitan Readiness Test (Oakland, 1972).

Gibb (1964) argues that individuals possess different amounts of TW, and Sarnacki (1979) asserts that students high in TW may more profitably employ their skills. He further contends that individuals low in TW are penalized and handicapped for lack of such abilities (Sarnacki, 1979). Since students differ in the amount of TW which they possess, it is also likely that different ethnic groups will possess varying amount of TW.

Major reasons for the need of this study are the following: (1) Ethnicity has not been used as an independent variable in the study of TW, and (2) the Tucson Unified School District (TUSD) is one of the few school districts in the United States

where large number of Blacks, Hispanics, Native-Americans, and Anglos can be tested in TW. If groups not skilled in TW are penalized as Sarnacki (1979) contends, and if minority groups possess different amounts of TW, then they are penalized and handicapped for the lack of these skills. Students above average in TW may be over-rewarded, too. Since standardized tests are used to evaluate student performance, and to assign students to varied education projects, then those students lacking in TW will be incorrectly evaluated and incorrectly assigned to such educational projects as a result of lack of TW skills.

Tests of TW have been developed by Gibb (1964), Millman (1966), Slakter et al., (1970a), Woodley (1973), and Bajtelsmit (1975a). These measures will be examined carefully in preparation for the present study.

If, indeed, students are penalized for lack of TW, or rewarded for possession of TW, and this ability is related to ethnic background, then there are ethnic groups penalized in the American educational system for lack of TW skills. Further, if it can be shown that TW can be taught to ethnic groups lacking TW, then deficiencies in TW can be remedied.

The general purpose of this project is to investigate the amount of TW in four ethnic populations (Black, Hispanic, Native-American, and Anglo) and how TW affects the reading comprehension scores of these four populations. This study will address the following research questions which focus on minority group populations: (1) Do Black, Hispanic, Native-American, and Anglo students possess different amounts of TW skills? (2) Will Black, Hispanic, Native-American, or Anglo students make greater gains from pretest to posttest on TW skills? (3) Will there be differences between treatment and control groups on reading comprehension at Grades 3, 5, and 7 after instruction in TW? (4) Will there be differences between treatment and control groups on testwiseness at Grades 3, 5, and 7 after instruction in TW? (5) Will Black, Hispanic, Native American, or Anglo students make greater gains in reading comprehension after acquiring TW skills? (6) Will third, fifth, and seventh grade students make different gains in acquiring TW skills?

METHOD

Sample Selection

Five junior high schools were selected from the Tucson Unified School District because they represented a cross-section of socioeconomic (SES) areas and their enrollment consisted of large number of minority groups. Thirteen elementary schools were selected for this study for the same reasons. At each school teachers were asked to voluntarily participate in this National Institute of Education funded study of TW. Although grades 3, 5, 7 and 8 were represented in this study, results are reported only for grades 3, 5 and 7. Refer to Table 1 for the names and grades of participating schools.

Table 1. Schools and Grades Participating in Testwiseness Study

School	Grade
A. Elementary Schools	
Cavett	3
Corbett	3,5
Davidson	3
Drachman	3,5
Holladay	5
Hollinger	3,5
Lawrence	3,5
Lynn	3,5
Pueblo Gardens	3,5
Rose	3,5
Safford	3,5
Van Buskirk	5
White	3,5
B. Junior High Schools	
Doolen	7
Naylor	7,8
Pistor	7
Safford	7,8
Wakefield	7,8

A total of 854 students (720 students in treatment groups; 134 students in control groups) participated in this study representing a total of 56 classrooms and 32 teachers. Further information on treatment and control groups is presented in Table 2. The combined grades consisted of 5 percent Native Americans, 7 percent Blacks, 53 percent Hispanics, and 35 percent Anglos. Males comprised 52 percent of the total group and females 48 percent.

Table 2. Treatment and Control Groups Participating in the Testwiseness Study*

Grade	Treatment Classes	Treatment Students	Control Classes	Control Students	Total Students
3	8	154	3	49	203
5	9	186	2	30	216
7	17	255	4	55	310
8	13	125	-	--	125
TOTAL	47	720	9	134	854

*Only students present for both the pretest and posttest with the Test of Testwiseness (TOT) are included in these figures.

Instrumentation

1. Test of Testwiseness (TOT): The TOT was adapted from a test of testwiseness developed by Slakter et al. (1970a). The adapted test contained 16 items measuring four major aspects of testwiseness. Only the total score was employed in the analysis. This adapted test was written simply so as to be understandable to students who were suspected of having limited reading ability. The same test was administered as a pre and posttest.

   The TOT was adapted for this study by the principal investigator and the consultant to the project from a previous test of testwiseness developed by Slakter and his associates (Slakter, Koehler, and Hampton, 1970b). In fact, the TOT so closely paralleled the Slakter et al. scale that it might be considered on elementary level of their test. Four aspects of TW were used in the construction of the TOT. These aspects were chosen from categories described by Millman et al. (1965). These same categories have been selected by other researchers because they were capable of being conveniently measured (Slakter et al., 1970). The four aspects selected from Millman et al. (1965) were:

   The examinee should be able to (1) Select the option which resembles an aspect of the stem, (2) Eliminate options which are known to be incorrect and to choose from among the remaining options, (3) Eliminate similar options, i.e., options which imply the correctness of each other, (4) Eliminate those options which include specific determiners.

   Items were constructed in such a way that the keyed responses could not be arrived at by knowledge of the subject matter. The final form of the scale consisted of 16 items with 4 items designed to measure each of the four categories selected from Millman et al. (1965). Since the TOT was intended to be administered to students in the third, fifth, and seventh grades, many of whom were suspected of low reading ability, the TOT was written with a simple vocabulary.

2. Reading Comprehension

   A. Third Grade: The California Achievement Test (CAT), Level 13, Reading Comprehension subtest was administered as a pretest (Form D) and as a posttest (Form C). This subtest consisted of 27 multiple choice items.

   B. Fifth Grade: The Comprehensive Tests of Basic Skills (CTBS), Form S, Level 2, Reading Comprehension Test was administered

as a pretest. The CAT, Form C, Level 15, Reading Comprehension Subtest, was administered as a posttest. The CTBS, Reading Comprehension Test consists of 45 items; The CAT Reading Comprehension subtest consists of 40 items.

C. Seventh Grade: The CAT, Form C, Level 17, Reading Comprehension subtest was administered as both a pretest and a posttest.

3. Socioeconomic Status (SES): SES was measured by participation in the Free Lunch Program for Low Income Children. These participants were coded with 1; the others were coded 0.

Procedure

During the week of November 10, 1980, five inservices were held in widely-separated areas of Tucson in order to provide small inservice meetings for teachers who volunteered to participate in the Testwiseness Study. These meetings did not last more than one hour. The following week, the week of November 17, five more inservices were held in various geographical areas of Tucson. Like the meetings the previous week, these meetings did not exceed one hour. Participating teachers attended one meeting the week of November 10 and one meeting November 17. Transparencies which presented TW principles were projected on a screen and discussed. An outline of each meeting and copies of the transparencies are presented in Appendix 1.

The emphases of the inservice meetings were the four elements of TW described by Millman, Bishop, and Ebel (1965):

Elements independent of test constructor or test purpose.

A. Time-using strategy.
   1. Begin to work as rapidly as possible with reasonable assurance of accuracy
   2. Set up a schedule for progress through the test.
   3. Omit or guess at items which resist a quick response.
   4. Mark omitted items, or items which could use further consideration, to assure easy relocation.
   5. Use time remaining after completion of the test to reconsider answers.

B. Error-avoidance strategy.
   1. Pay careful attention to directions, determing clearly the nature of the task and the intended basis for response.
   2. Pay careful attention to the items, determing clearly the nature of the question.
   3. Ask examiner for clarification when necessary, if it is permitted.
   4. Check all answers.

C. Guessing strategy.
   1. Always guess if right answers only are scored.
   2. Always guess if the correction for guessing is less severe than a "correction for guessing" formula that gives an expected score of zero for random responding.
   3. Always guess even if the usual correction or a more severe penalty for guessing is employed, whenever elimination of options provides sufficient chance of profiting.

D. Deductive reasoning strategy.
   1. Eliminate options which are known to be incorrect and choose from among the remaining options.
   2. Choose neither or both of two options which imply the corrections of each other.
   3. Choose neither or one (but not both) of two statements, one of which, if correct, would imply the incorrectness of the other.
   4. Restrict choice to those options which encompass all of two or more given statements known to be correct.
   5. Utilize relevant content information in other test items and options.

Testwiseness instruction began approximately December 1, 1980 and ended around March 20, 1981. It was planned that each teacher would spend a minimum of one 15-minute session every two weeks teaching and practicing TW with their students. This would total to seven classroom sessions of 15-minutes each. One month after the end of the TW instruction, students were administered the California Achievement Test on April 21-24, 1981 as part of a statewide achievement testing program. The Reading Comprehension subtest was the posttest for the present TW study.

RESULTS

An investigation of the reliability and validity of the TOT was considered crucial to inferences drawn about TW and its effects in the present study. Therefore, a validity study was conducted on the TOT to investigate internal consistency, stability, convergent validity, discriminant validity and sensitivity to instruction.

The TOT initially consisted of 10 items designed to measure Risktaking and 16 items to measure Testwiseness. Five items of the Risktaking scale focused on risktaking while the other 5 items were filler items. With a total possible score on risktaking of 5, the analysis reveals most of the students obtained the total score and in many classrooms there was no variance. Therefore, the principal investigator and the consultant decided to eliminate risktaking from any analysis and focus on the 16-item TW scale.

The internal consistency of the TOT was estimated with Kuder-Richardson Formula 20 (KR20) reliability coefficients computed separately at each grade using pretest results. Next, KR20 coefficients were calculated separately for experimental and control groups at each grade level using posttest results. This was to avoid any student-treatment interaction resulting from the fact that some of the students were in experimental groups and others were in control groups. The median KR20 coefficient across grades and testings was .60. Stability coefficients over a five month period ranged from .36 to .49. Further analyses indicated the TOT was sensitive to instruction, and the evidence of convergent validity was satisfactory. Discriminant validity coefficients appeared as low as should be expected. A complete report of the TOT validity study appears in Appendix 2 and a copy of the TOT is in Appendix 3. The same TOT items were administered to third, fifth and seventh grade students participating in the TW study. In summary, the TOT was deemed comparable to previous TW scales, and its reliability and validity were considered satisfactory. Only the total test score would be used in analyses.

A preliminary analysis was conducted for each classroom participating in the present study in order to provide results to the participating teachers (Refer to Appendix 4). Often teachers remark that researchers do not inform them of the results of studies which are conducted with the cooperation of classroom teachers. For that reason, a special effort was made to contact all participating teachers. The findings of this

preliminary analysis, involving pre- to posttest comparisons of dependent samples using t-test with each classroom, indicate the following:

1. Testwiseness can be taught.
2. Testwiseness can be taught as early as the third grade.
3. A test developed to measure testwiseness will detect changes resulting from testwiseness instruction.
4. Students will not gain in test-taking skills simply by maturation or development of other skills. However, specific instruction in testwiseness can raise testwiseness ability.

The following are the six research questions addressed by the present research study:

1. Do Black, Hispanic, Native American and Anglo students possess different amounts of TW skills?

2. Will Black, Hispanic, Native American or Anglo students make greater gains from pretest to posttest on TW skills?

3. Will there be differences between treatment and control groups on reading comprehension at Grades 3, 5, and 7 after instruction in TW?

4. Will there be differences between treatment and control groups on testwiseness at Grades 3, 5 and 7 after instruction in TW?

5. Will Black, Hispanic, Native American, or Anglo students make greater gains in reading comprehension after instruction in TW?

6. Will third, fifth and seventh grade students make different gains in acquiring TW skills?

RESEARCH QUESTION 1: *Do Black, Hispanic, Native American, and Anglo students possess different amounts of TW skills?*

The four ethnic groups were compared on their pretest TOT scores statistically controlling for different pretest reading levels and SES. The main effects for ethnicity in the third grade were not significant $F(3,121)=.72$, $p<.539$; nor were the main effects

for ethnicity significant at the fifth grade, $F(3,193)=1.64$, $p< .183$; nor at the seventh grade, $F(3,261)=1.36$, $p< .255$. The results were consistent for all three grades. Therefore, it was concluded that ethnic groups were not different in their testwiseness skills before testwiseness instruction.

Because of the significant variance accounted for by reading ability, the use of reading as a covariate was justified. SES, on the other hand, was only significant at the third grade, $p< .027$. Since SES was not significant at the fifth and seventh grades, it could have been dropped from the analysis in order to gain greater power in comparing the ethnic groups. Since the influence of SES has been documented in a large body of research, SES was allowed to remain in the analyses. Refer to Tables 3-5 for greater detail on the analysis.

Table 3. Analysis of Covariance on Third Grade Pretest Test of Testwiseness (TOT) Scores Comparing Ethnic Groups

A. Summary Table

Source of Variation	df	MS	F	p
Covariates:				
Pretest (Reading)	1	90.49	27.09	.000
SES	1	16.71	5.00	.027
Main Effects:				
Ethnicity	3	2.42	.72	.539
Residual	121	3.34		

B. Adjusted and Unadjusted Means

Group	N	Unadjusted Means	Adjusted-Means
Native American	7	7.85	8.99
Black	15	8.66	9.03
Hispanic	63	8.92	8.93
Anglo	42	9.83	8.77

Grand Mean = 9.13

Table 4. Analysis of Covariance on Fifth Grade Pretest Test of Testwiseness (TOT) Scores Comparing Ethnic Groups

A. Summary Table

Source of Variation	df	MS	F	p
Covariates:				
Pretest (Reading)	1	207.22	44.19	.000
SES	1	4.06	.86	.354
Main Effects:				
Ethnicity	3	7.67	1.64	.183
Residual	193	4.69		

B. Adjusted and Unadjusted Means

Group	N	Unadjusted Means	Adjusted Means
Native American	4	10.25	10.60
Black	9	7.45	8.54
Hispanic	106	9.65	9.87
Anglo	80	9.86	9.43

Grand Mean = 9.65

Table 5. Analysis of Covariance on Seventh Grade Pretest Test of Testwiseness (TOT) Scores Comparing Ethnic Groups

A. Summary Table

Source of Variation	df	MS	F	p
Covariates:				
Pretest (Reading)	1	64.58	18.43	.000
SES	1	.21	.06	.809
Main Effects:				
Ethnicity	3	4.77	1.36	.255
Residual	261	3.50		

B. Adjusted and Unadjusted Means

Group	N	Unadjusted Means	Adjusted Means
Native American	9	9.67	9.89
Black	14	10.86	11.02
Hispanic	145	9.83	9.98
Anglo	99	10.38	10.12

Grand Mean = 10.08

RESEARCH QUESTION 2: *Will Black, Hispanic, Native American or Anglo students make greater gains from pretest to posttest on TW skills?*

A three factor repeated measures ANOVA (Ethnicity X SES X Time) was computed for third, fifth, and seventh grade students. The time factor was represented by pre- and posttesting with the TOT. Differential gains for some ethnic groups would be reflected in a time by ethnicity interaction. Such an interaction did not appear in the analyses. For the third grade the T X E interaction was non-significant, $F(3,152)=.37$, $p < .772$, nor was this two factor interaction significant at the fifth grade, $F(3,182)=1.05$, $p < .372$, nor at the seventh grade, $F(3,216)=.56$, $p < .640$. Therefore, it was concluded the gains of each ethnic group were comparable across time. Different ethnic groups did not appear to make different gains in TW. These results are presented in Tables 6-8 and the mean TOT scores for each group are displayed in Tables 9-11.

Table 6. Three Factor Repeated Measures ANOVA on Testwiseness with Third Grade Students

Source of Variation	df	MS	F	p
Between subjects				
Ethnicity (E)	3	22.57	2.91	.036
SES (S)	1	36.98	4.77	.030
ES	3	6.28	.81	.490
Subj w. groups	152	7.75		
Within subjects				
Testwiseness (T)	1	173.32	62.87	.000
TE	3	1.03	.37	.772
TS	1	.41	.15	.701
TES	3	6.05	2.20	.091
TX subj w. groups	152	2.76		

TABLE 7: Three Factor Repeated Measures ANOVA on Testwiseness with Fifth Grade Students

Source of Variation	df	MS	F	p
Between subjects				
Ethnicity (E)	3	15.07	1.56	.201
SES (S)	1	.03	.00	.953
ES	3	2.62	.27	.846
Subj w. groups	182	9.66		
Within subjects				
Testwiseness (T)	1	41.91	11.90	.001
TE	3	3.70	1.05	.372
TS	1	1.32	.38	.540
TES	3	.25	.07	.976
Subj w. groups	182	3.52		

TABLE 8: Three Factor Repeated Measures ANOVA on Testwiseness with Seventh Grade Students

Source of Variation	df	MS	F	p
Between subjects				
Ethnicity (E)	3	27.65	3.19	.024
SES (S)	1	.04	.00	.947
ES	3	8.39	.97	.408
Subj w. groups	216	8.67		
Within subjects				
Testwiseness (T)	1	127.44	14.89	.000
TE	3	4.81	.56	.640
TS	1	16.36	1.91	.168
TES	3	5.32	.62	.601
Tx subj w. groups	216	8.56		

Table 9: Descriptive Statistics of Third Grade Students by Ethnicity and Time (Pretest/Posttest) on Testwiseness

Ethnicity	Pretest			Posttest		Difference	Total
	N	Mean	SD	Mean	SD	Means	Mean
Native-American	9	8.56	3.00	10.78	3.33	2.33	9.72
Black	21	8.76	2.26	11.52	2.40	2.76	10.14
Hispanic	77	8.90	2.12	10.60	2.60	1.70	9.75
Anglo	53	9.96	1.85	12.07	2.27	2.11	11.02
TOTAL	160	9.21		11.22		2.01	10.22

Table 10: Descriptive Statistics of Fifth Grade Students by Ethnicity and Time (Pretest/Posttest) on Testwiseness

Ethnicity	Pretest			Posttest		Difference	Total
	N	Mean	SD	Mean	SD	Means	Mean
Native-American	4	10.50	1.29	11.25	1.71	.75	10.87
Black	6	8.17	2.40	10.67	2.07	2.50	9.42
Hispanic	106	9.55	2.48	10.67	2.79	1.12	10.11
Anglo	74	9.91	2.40	11.80	2.53	1.89	10.85
TOTAL	190	9.66		11.12		1.46	10.39

Table 11: Descriptive Statistics of Seventh Grade Students by Ethnicity and Time (Pretest/Posttest) on Testwiseness

Ethnicity	Pretest			Posttest		Difference	Total
	N	Mean	SD	Mean	SD	Means	Mean
Native-American	11	10.00	2.49	11.73	1.90	1.73	10.86
Black	14	10.86	1.92	11.57	2.90	.71	11.21
Hispanic	122	9.63	2.45	12.28	4.24	2.65	10.95
Anglo	77	10.62	1.83	13.14	2.19	2.52	11.88
TOTAL	224	10.07		12.50		2.43	11.29

RESEARCH QUESTION 3: Will there be differences between treatment and control groups on reading comprehension at Grades 3, 5, and 7 after instruction in TW?

Students' scores on all Reading Comprehension tests (CTBS or CAT, Forms C or D) were converted to T-scores before analysis to make test scores comparable. Treatment and control groups were compared to determine if there were differences due to TW instruction. Control groups were established at Safford and Pueblo Gardens elementary schools. These control groups were compared with treatment groups at Cavett, Hollinger, Pueblo Gardens and Safford Elementary schools. These treatment schools were selected because the percentages of students in low income categories were similar based on ESEA Title I school rankings. Control groups were also formed at Pistor and Wakefield Junior High Schools. These groups' were compared with treatment groups at Pistor, Wakefield and Safford Junior High Schools.

Treatment and control groups were compared on reading comprehension posttest scores adjusting for differences on reading pretest scores. There were no significant differences between treatment and control groups at the third grade, $F(1,37)=.025$, $p< .876$; at the fifth grade, $F(1,63)=1.39$, $p< .243$; and at the seventh grade, $F(1,168)=1.67$, $p< .198$. These results suggest that there were no differences between treatment and control groups due to testwiseness training. Refer to Tables 12-14.

Table 12. Analysis of Covariance on Third Grade Reading Comprehension Posttest Scores Comparing Treatment and Control Groups.

A. Summary Table

Source of Variation	df	MS	F	p
Covariate: Pretest (Reading)	1	293.18	59.78	.000
Main Effects: Group	1	.12	.025	.876
Residual	37	4.904		

B. Adjusted and Unadjusted Means

Group	N	Unadjusted Means	Adjusted Means
Treatment	27	35.48	35.82
Control	13	36.41	35.70

Grand Mean = 35.78

Table 13. Analysis of Covariance on Fifth Grade Reading Comprehension Posttest Scores Comparing Treatment and Control Groups

A. Summary Table

Source of Variation	df	MS	F	p
Covariate:				
Pretest (Reading)	1	2331.84	62.09	.000
Main Effects:				
Group	1	52.12	1.39	.243
Residual	63	37.56		

B. Adjusted and Unadjusted Means

Group	N	Unadjusted Means	Adjusted Means
Treatment	55	43.57	43.19
Control	11	43.68	45.60

Grand Mean = 43.59

Table 14. Analysis of Covariance on Seventh Grade Reading Comprehension Posttest Scores Comparing Treatment and Control Groups.

A. Summary Table

Source of Variation	df	MS	F	p
Covariate:				
Pretest (Reading)	1	5341.19	139.52	.000
Main Effects:				
Group	1	63.88	1.67	.198
Residual	168	38.28		

B. Adjusted and Unadjusted Means

Group	N	Unadjusted Means	Adjusted Means
Treatment	128	47.20	47.23
Control	43	45.93	45.83

Grand Mean = 46.88

RESEARCH QUESTION 4: Will there be differences between treatment and control groups on testwiseness at Grades 3, 5 and 7 after instruction in TW?

Analyses of covariance were used to compare treatment and control groups on testwiseness posttest scores adjusting for pretest TW differences. These comparisons included the same schools as did the comparisons for Research Question 4. However, there are differences in sample sizes because more students were pre- and posttested with the TOT. ANCOVA results indicated significant differences between treatment and control groups favoring the treatment groups in the third grade, $F(1,85)=11.48$, $p< .001$; the fifth grade, $F(1,63)=8.39$, $p< .005$; and the seventh grade, $F(1,168)=24.30$, $p< .001$. These results indicate treatment and control groups were different on the posttest. These findings are further supported by the preliminary class by class analysis (Appendix 4) in which TW classrooms consistently showed significant gains whereas control groups did not show significant gains. Refer to Tables 15-17.

Table 15. Analysis of Covariance on Third Grade Testwiseness Posttest Scores Comparing Treatment and Control Groups

A. Summary Table

Source of Variation	df	MS	F	p
Covariate: Pretest (Testwiseness)	1	107.54	17.78	.000
Main Effects: Group	1	69.39	11.48	.001
Residual	85	6.05		

B. Adjusted and Unadjusted Means

Group	N	Unadjusted Means	Adjusted Means
Treatment	58	10.75	10.86
Control	30	8.96	8.97

Grand Mean = 10.22

Table 16. Analysis of Covariance on Fifth Grade Testwiseness Posttest Scores Comparing Treatment and Control Groups

A. Summary Table

Source of Variation	df	MS	F	p
Covariate:				
Pretest (Testwiseness)	1	54.52	11.02	.002
Main Effects:				
Group	1	41.51	8.39	.005
Residual	63	4.95		

B. Adjusted and Unadjusted Means

Group	N	Unadjusted Means	Adjusted Means
Treatment	55	10.83	10.76
Control	11	8.18	8.56

Grand Mean = 10.39

Table 17. Analysis of Covariance on Seventh Grade Testwiseness Scores Comparing Treatment and Control Groups

A. Summary Table

Source of Variation	df	MS	F	p
Covariate:				
Pretest (Testwiseness)	1	223.72	42.10	.000
Main Effects:				
Group	1	129.14	24.30	.000
Residual	168	5.313		

B. Adjusted and Unadjusted Means

Group	N	Unadjusted Means	Adjusted Means
Treatment	128	12.13	12.08
Control	43	9.95	10.08

Grand Mean = 11.58

RESEARCH QUESTION 5: Will Black, Hispanic, Native American or Anglo students make greater gains in reading comprehension after instruction in TW?

The results of Research Question 3 presented evidences that treatment and control groups were not different in reading comprehension after instruction in TW. This indicated that TW instruction did not have an effect on standardized reading comprehension test scores. If pre- to posttest reading comprehension differences could not be attributed to TW instruction, then Research Question 5 was no longer a logical part of the present study, and it is not presented. For those who are curious about the results of the analysis of Research Question 5, reference is made to Appendix 5.

RESEARCH QUESTION 6: Will third, fifth and seventh grade students make different gains in acquiring TW skills?

The mean gain of third grade students in testwiseness was 2.01, of fifth grade students was 1.46 and of seventh grade students was 2.43. There was not sufficient evidence to consider these gains significantly different. Refer to Tables 8-10.

DISCUSSION

The present study of TW is important because it was the first study of TW to focus on ethnic group differences. This study included four major ethnic groups: Native American, Black, Hispanic and Anglo. Furthermore, Reading Comprehension and Socioeconomic Status, which were selected as control variables, were most crucial in the study of TW because research has shown that these three variables are highly intercorrelated and easily confounded.

The logic of this study was based on two premises from which inferences were drawn about possible ethnic group differences: (1) The Test of Testwiseness (TOT) was an adequate measure of testwiseness, and (2) the testwiseness instruction was effective. Evidence from the validation analysis of the TOT indicated the TOT was adequate. Treatment and control group comparisons demonstrated the testwiseness instruction was effective.

The results of the present study did not show the effects of TW on Reading Comprehension scores of the CAT. Treatment and control groups were not significantly different. Several factors may account for such a finding:

(1) Each teacher received only two hours of TW instruction and they provided a minimum of one hour and forty-five minutes total time across four months, from December through March, in the instruction of TW. There simply may not have been enough instruction of the teachers, and the teachers may not have spent enough time instructing the students.

(2) The focus of instruction was on the following aspects of TW: (a) stem-option resemblance, (b) elimination of incorrect or absurd options, (c) elimination of similar options, and (d) elimination of options with specific determiners. These aspects of TW appeared to have been removed from the California Achievement Test. Thus, it may be inferred that test specialists who develop standardized tests may identify and delete such TW clues during the test construction process.

Two salient findings resulted from the present research: Before instruction in TW, the ethnic groups were not different in the amount of TW they possessed when differential reading abilities and SES were

controlled. This was found at the third, fifth and seventh grades. This finding was particularly convincing because different reading abilities and SES levels were controlled. Unless reading and SES are controlled, these variables will confound any ethnic group comparisons because reading abilities of minority students are often lower than that of Anglo students, and minority students often are of a lower SES level than Anglo students. In fact, in the analysis of Research Question 1, it was found the four ethnic groups were significantly different in their reading ability. If reading ability were not controlled, one might discover significant differences among ethnic groups on TW and not realize that the difference is due to Reading Comprehension or SES.

Another finding is that when instructed in TW, ethnic groups made comparable gains from pretest to posttest. Such comparable gains are rarely found in achievement testing in reading and mathematics. Thus, not only does this research find evidence that ethnic groups possess comparable TW skills, but it also finds that when instructed in TW, different ethnic groups appear to make comparable gains in the acquisition of TW skills.

One purpose of the present study was to investigate the amount of TW in four ethnic groups. It has been argued that if ethnic groups possess different amounts of TW, then those with less amounts of TW will be penalized in the American educational system. It has also been argued that if there are deficiencies, then any penalizing can be overcome through instruction. The present study has shown that different ethnic groups do not possess significantly different amounts of TW, and when they are instructed in TW, they make comparable gains.

APPENDIX 1

Testwiseness Instruction

National Institute of Education

The Effect of Testwiseness on the Reading Achievement Scores of Minority Populations

Week of November 10, 1980

3:15 p.m. to 4:15 p.m.

Principal Investigator: Dr. Stephen Powers Phone -6138
Project Consultant: Dr. Darrell Sabers

Outline

1. Schedule of Sessions
2. Teacher-Class Information
3. Overview of the Project
4. What is testwiseness?
5. Testwiseness Pretest: Critique
6. Control Groups
7. Third Grade Teachers: Third Grade Reading Test

Next week: Methods of Teaching Testwiseness

MEMO

Department of
LEGAL and RESEARCH SERVICES
Division of PLANNING, ANALYSIS and MANAGEMENT

TO: Participating Teachers in the NIE Testwiseness Study

FROM: Stephen Powers, Principal Investigator
NIE Testwisness Study, Legal and Research Services

SUBJECT: National Institute of Education's Testwiseness Study

Thank you for your volunteering to participate in the NIE Testwiseness Study. Two one-hour meetings are planned after school during the weeks of November 10 and November 17. Each week you may choose one meeting to attend at any of the five schools. All the presentations on testwiseness during the week of November 10 will be identical. The same will be true of the presentations during the week of November 17. The locations and times of presentations on testwiseness are presented below. The numbers of participants at each session are expected to be between 5 and 15.

Week of November 10: What is testwiseness?

Day	Mon.	Tues.	Wed.	Thurs.	Fri.
Date	10	11	12	13	14
School	Doolen	Naylor	Safford	Wakefield	Pistor
Time	3:15	3:15	3:15	3:15	3:15
Room	44	79	Library	Library on West side	Science Room

Week of November 17: Methods of teaching testwiseness

Day	Mon.	Tues.	Wed.	Thurs.	Fri.
Date	17	18	19	20	21
School	Safford	Wakefield	Pistor	Doolen	Naylor
Time	3:15	3:15	3:15	3:15	3:15
Room	Library	Library on West side	Science Room	44	79

Dr. Darrell Sabers, the project's consultant, and I are looking forward to meeting you at these two sessions. We have prepared a packet of materials on testwiseness and testtaking skills which will be presented to you as professional reference material. Moreover, we have adapted the essentials of testwiseness for a presentation to you.

For your interest, I have attached a list of participating schools.

If I can be of any assistance, please call me at -6138.

Steve Powers

SP/ch
10/24/80

Attachment

Copies to Principals
Darrell Sabers

NIE Testwiseness Study 1980-81
Participating Schools

Junior High Schools	Elementary Schools
Doolen	Cavett
Naylor	Corbett
Pistor	Davidson
Safford	Drachman
Wakefield	Holladay
	Hollinger
	Lawrence
	Lynn
	Pueblo Gardens
	Rose
	Safford
	Van Buskirk
	White

TUSD L&R
SP/ch
10/24/80

NIE STUDY OF TESTWISENESS

TEACHER-CLASS INFORMATION FORM

1. NAME _____

2. SCHOOL _____

3. ELEMENTARY SCHOOL TEACHERS: Approximately, how many students are in your class who will receive testwiseness instruction?

_____

4. JUNIOR HIGH SCHOOL TEACHERS: In what class periods will you give testwiseness instruction? Approximately, how many students are in each class? What subject is taught in each class? Fill in your answers below:

Period	Subject	Approximate number of students
___	_____	_____
___	_____	_____
___	_____	_____
___	_____	_____
___	_____	_____
___	_____	_____

5. SPECIAL ASSIGNMENT TEACHERS: If you are a teacher of learning disability students, gifted students, counselors, or other responsibilities, please state what your educational assignment is.

Thank you

Steve Powers
11/10/80

OUTLINE OF TESTWISENESS INSTRUCTION

1. Research Funded by NIE: The Effect of Testwiseness on the Reading Achievement Scores of Minority Population. Many people interested.

   Describe the Research (including control group and SES)

2. Hand out TW Tests and Information Sheets

3. Theory

   Definition: "a subject's capactiy to utilize the characteristics and formats of the test and/or the test-taking situation to receive a high score" - Millman, Bishop and Ebel (1965)

   A) TW is independent of the examiner's knowledge of the subject.

   B) Correlated with testtaking experience (limited amount of practice only).

   C) Negatively correlated with time (eroded with time). Adults appeared deficient in TW.

   D) Moderate positive relationship with intelligence.

   E) High correlation with verbal achievement.

   F) Negative correlation with test anxiety.

   G) No relationship with sex

   H) Correlation with grade level. Linear trend

4. Research

   A) Can be measured

   B) Can be taught

   C) Training in TW skills improved performance on TW tests

   D) Training in TW skills improve performance on teacher tests

   E) Training in TW skills improves performance on standardized tests

   F) Students high in TW more profitably use their skills

   G) Students low are penalized

      Students scored higher when:

      1) Time limit removed

      2) Received verbal praise and encouragement

      3) Tests labeled a game and in a game group

Students scored higher when: (contd.)

4) An IQ test instead of achievement test

5) Reinforced for correct answers

6) When instructions are clear

Test-Taking Skills

1) Teacher influence: prepare room, explain clearly, etc.

2) General Strategies

A) Answer ALL questions

B) Don't be a quitter

C) Follow directions

D) Don't look for a pattern

E) Ask questions if you don't understand

F) Use time wisely

3) Testwiseness Skills

A) Stem option resemblance

B) Eliminate similar options

C) Eliminate absurd option, then answer to the best of your ability, the remaining options

D) Avoid specific determiners

TUSD L&RS
SP/se
2/18/82

NIE Testwiseness Study

Week of November 17, 1980

3:15 p.m. to 4:15 p.m.

AGENDA

1. Review of TW Principles (slides)
2. Overview
3. Testwiseness
4. NIE Testwiseness Study
5. Practice Items
6. Standardized Tests
7. Control Groups
8. Discuss the TW test which will be sent to teachers
9. Discuss methods of teaching TW
10. Tips on Discussing Testwiseness (TW) with children

SP/ch
11/17/80

NIE Testwiseness Project

OVERVIEW

The testwiseness instruction begins approximately December 1, 1980 and ends around March 20, 1981. It is hoped that participating teachers will spend a minimum of one 15-minute session every two weeks discussing and practicing testwiseness skills every two weeks while school is in session. This would mean one session in December, two sessions in January, two sessions in February and two sessions in March. This will make a total of seven instructional sessions. If participating teachers wish to give longer instruction in testwiseness they may.

Statewide achievement testing of students in Grades 1-12 in reading, mathematics and language arts will be conducted in Tucson Unified School District April 21-24, 1981. It is hoped that the teachers participating in the testwiseness study have given their students special preparation for the statewide testing so that they may apply many of these test-taking principles.

SP/ch
11/17/80

Testwiseness

Testwiseness is "a subject's capacity to utilize the characteristics and formats of the test and/or the test-taking situation to receive a high score."

The following are some of the items we have already discussed:

1. Principle: Avoid Specific Determiners

   The Dakstan desert is

   A. always hot in March.
   B. never windy in May.
   C. sometimes temperate
   D. only cold in December.

2. Principle: Eliminate Similar Options

   Bendaline's opera "The Three Bells" is about

   A. marital problems.
   B. domestic strife.
   C. the problem of aging.
   D. the conflict between an old man and his wife.

3. Principle: Stem-Option Resemblance

   Which of the following instruments measures self-esteem?

   A. The Campbell Interest Test
   B. The Kidder Preference Inventory
   C. The Vassey Intelligence Scale
   D. The Utah Self-Concept Scale

4. Principle: Eliminate Absurd Options

   America's most popular sport is

   A. sewing.
   B. baseball.
   C. reading.
   D. knitting.

NIE TESTWISENESS STUDY

1. Teacher Influence: Prepare room, explain clearly, etc.

2. General Strategies

   a. Answer All Questions
   b. Don't be a quitter
   c. Follow Directions
   d. Don't look for a pattern
   e. Ask questions if you don't understand
   f. Use Time Wisely

3. Testwiseness Skills

   a. Stem option resemblance
   b. Eliminate similar options
   c. Eliminate absurd options, then answer to the best of your ability, the remaining options
   d. Avoid specific determiners

MEMO

Department of
LEGAL and RESEARCH SERVICES
Division of PLANNING, ANALYSIS and MANAGEMENT

TO: Teachers in the Testwiseness Study

FROM: Stephen Powers, Principal Investigator
NIE Testwiseness Project

SUBJECT: Practice Testwiseness Items

Enclosed are the following:

A. Practice items for teaching your students many testwiseness skills. These items are categorized as

1. Stem-Option Resemblance,
2. Specific Determiners,
3. Absurd Options, and
4. Similar Options.

The grade levels for these items vary from third to seventh grade. Many of these items you will want to reword to make them more relevant to your students ability level. Feel free to make any changes you feel are needed.

B. A summary of the testwiseness principles.

C. A summary of testwiseness strategies and skills.

If I can be of any assistance, please call me at -6138.

SP/ch
1/7/81

Stephen Powers

The correct response is starred. The similar words are circled.

1. Stem-Option Resemblances:

A (friend) in need is

1. John's brother
2. the banker
3. a policeman
*4. a (friend) indeed

The story "The (Tortoise) and The Hare" is about

1. racing
*2. a (tortoise) who wins a race
3. rabbits
4. animal crackers

(Cactus) wrens build their nests in

A. Bird houses
B. mesquite trees
*C. Sahuaro (cactus)

Which of the following is a (dinosaur)?

*A. Bronto(saurus)
B. Elephant
C. Pithicantropus
D. Neanderthal

The (fire) ant can be identified by its

A. large pincers
B. armored appearance
C. long legs
*D. (bright red color)

The short story, "Life in An (Eskimo) Village" was written by

A. a person who traveled in the Northwest
B. a professor who teaches mathematics
C. a third grader at Smith Elementary School
*D. an (Eskimo) who lived in Alaska

The story ("A Trip") is about

A. George Washington's childhood
*B. (a train ride)
C. dinosaurs
D. the bakery

1. Stem-Option Resemblances (contd.):

To bother the neighbors means to

A. insult your friends
*B. disturb the neighbors
C. disrupt the game
D. shout in the woods

The proper way is the

*A. correct way
B. easiest answer
C. best road
D. fastest player

He intended to work means that he had

A. taken a job
B. planned a vacation
*C. decided to work
D. started working

In order to change an answer you should

A. mark another response
B. make two marks
C. erase all marks
*D. change your mark

The mountain cabin is

*1. on Mount Rushmore
2. on the low desert
3. in the valley
4. on the Great Plains

The correct answers are starred. The specific determiners are underlined.

2. Specific Determiners:

A wicked king is

A. always rich
*B. sometimes kind
C. never wise
D. always happy

A fast boat

1. always wins a race
*2. sometimes is big
3. is never small
4. only has one driver

Billy's best friend

1. never argues with him
2. always agrees with him
*3. is often late
4. only likes candy

Japanese children

A. always wear uniforms
B. never eat beef
C. eat rice only at dinner
*D. sometimes go on field trips

Seeds are different from eggs because

A. they always grow in the ground
B. their husks can never be eaten
*C. they are produced from plants instead of animals
D. all of them can be eaten

People who eat well

A. always get fat
B. spend all of their time cooking
*C. usually don't need vitamins
D. never get sick

Learning how to drive at night is

A. never as easy as learning in the daytime
B. always a strain on the eyes
*C. sometimes harder than driving in the daylight
D. only useful if the person plans to drive at night

2. Specific Determiners (contd.):

A modern computer is

*A. sometimes useful in solving problems
B. always reliable
C. the only way to gather information
D. used only by computer scientists

An airplane mechanic

1. never flies in a plane
*2. sometimes travels by plane
3. only travels by car
4. always likes his work

A wild horse

1. never can be tamed
2. always runs with other horses
3. only rests at night
4. none of the above

The student should learn to respond to "none of the above"

The correct options are starred.

3. Absurd Options:

	NOTES
Tom's pet Herman is 1. an elephant 2. a clown *3. a tortoise 4. a go-cart	Choices 1, 2, 4 are absurd
Hot means *A. warm B. cold C. sweet D. rough	Choices B, C, d are absurd
Six minus two equals A. eight B. twelve *C. four D. twenty-four	This problem mathematically is 6 - 2 = ? The choices A, B, D are absurd because they are much larger than 6.
A deer is *A. an animal B. a house C. a barn D. a cat	Choices B, C, D are absurd
To go means *A. to leave B. to study C. to laugh D. to work	Choices B, C, D are absurd
Where was the first United States Mint? A. in Germany B. in Berlin C. in Europe *D. in Philadelphia	Choices A, B, C are absurd because they are not in the U.S.
If you water your grass, A. it will die *B. it will grow C. it burns D. the weeds die	Choices A, C, D are absurd

3. Absurd Options (contd.):

NOTES

The man with many cats

A. is a workman
B. likes to study
C. is a good friend of John
*D. is an animal lover

Choices A, B, C are unrelated or absurd

A silly idea is

*A. foolish
B. good
C. strange
D. new.

Choices B, C, D are absurd

Where is the Statue of Liberty?

*A. in the United States
B. in France
C. in Spain
D. in Russia

Choices B, C, D are absurd because they are not in the U.S.

Alaska is a bigger state than Texas because

A. it was added to the U.S. at a later time
*B. there are more square miles of it
C. there are more people in it
D. there is a greater distance between cities

Choices A, C, D are absurd

The pioneers traveled by

A. Toyota
B. plane
C. bus
*D. covered wagon

Choices A, B, C are absurd

The correct options are starred.

4. Similar Options:

NOTES

To invite a friend means

1. fast
2. quick
3. speedy
*4. ask

Choices 1, 2, 3 are similar

During the storm the lake was

1. silent
2. quiet
3. still
4. peaceful
*5. turbulent

Similar Options--the student should be able to eliminate four options and get the answer correct without knowing the word "turbulent."

Where did Joe's mother take him to ride

A. into the city
B. down town
C. to main street
*D. in the country

A, B, C are similar

Darin's pet is a

1. tortoise
2. turtle
*3. puppy

1, 2 are similar

Where is the Golden Gate Bridge?

A. in the United States
B. in New York
C. In New York State
*D. in San Francisco

A, B, C are similar

To open a Gilt box you should hold the handle and

1. twist it
2. turn it
*3. pull it

1, 2 are similar

How did Bill cross the street?

1. he dashed across
2. he rushed across
3. he hurried across
*4. he walked slowly

1, 2, 3 are similar

4. Similar Options (contd.):

NOTES

Most reptiles have

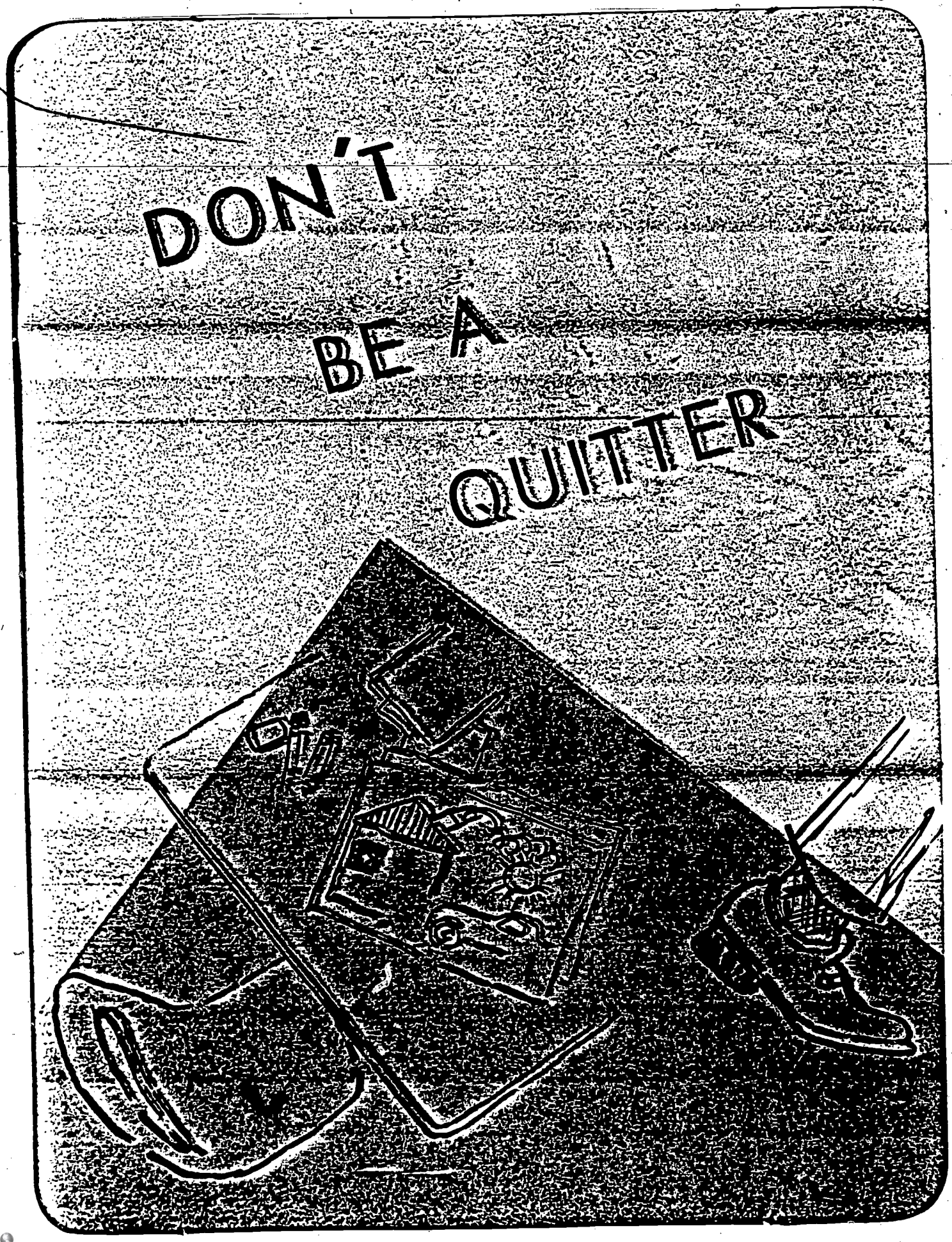
A. bright skins
B. many colors on their skins
C. multicolored features
*D. creep or crawl

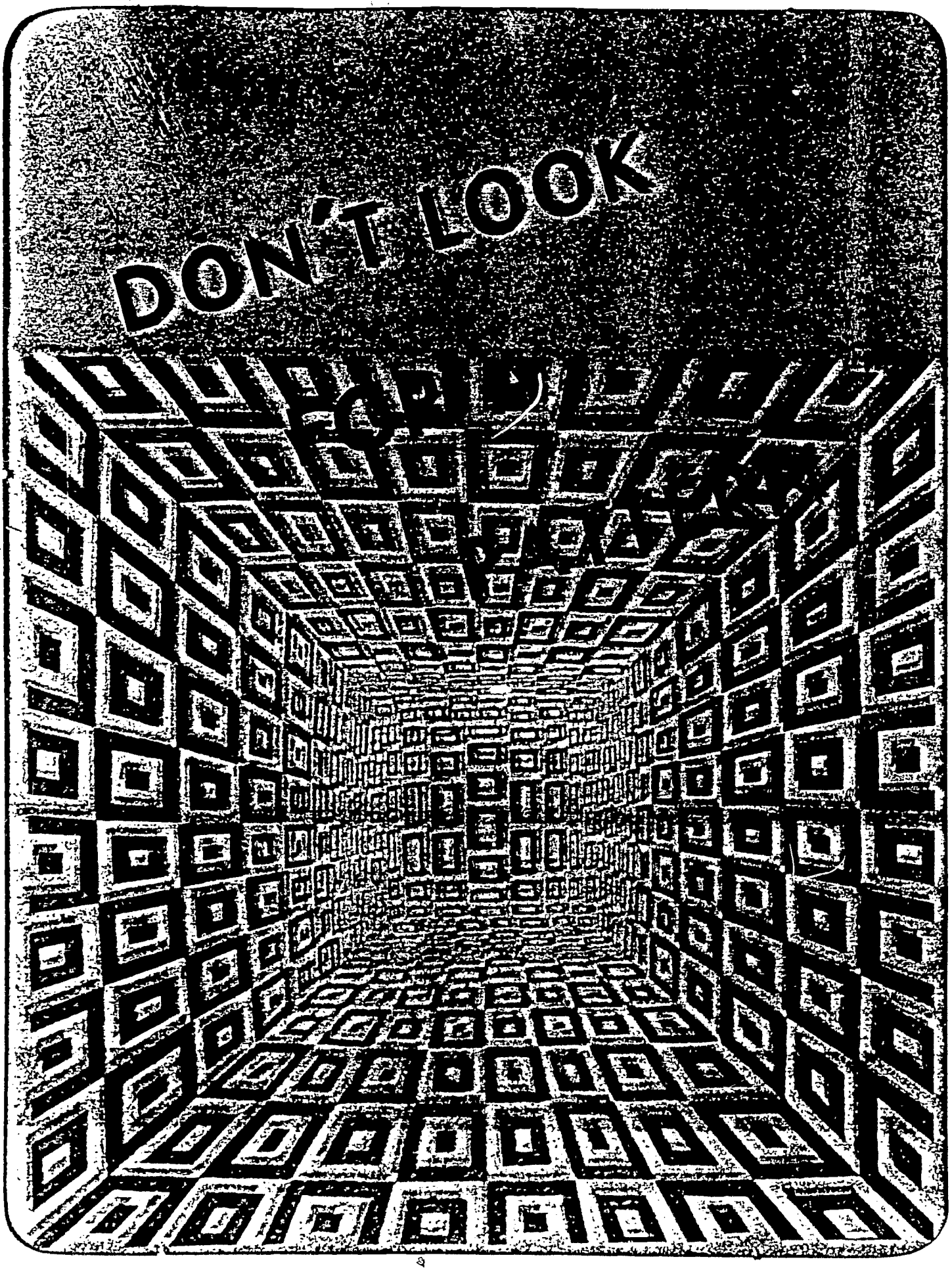
A, B, C are similar

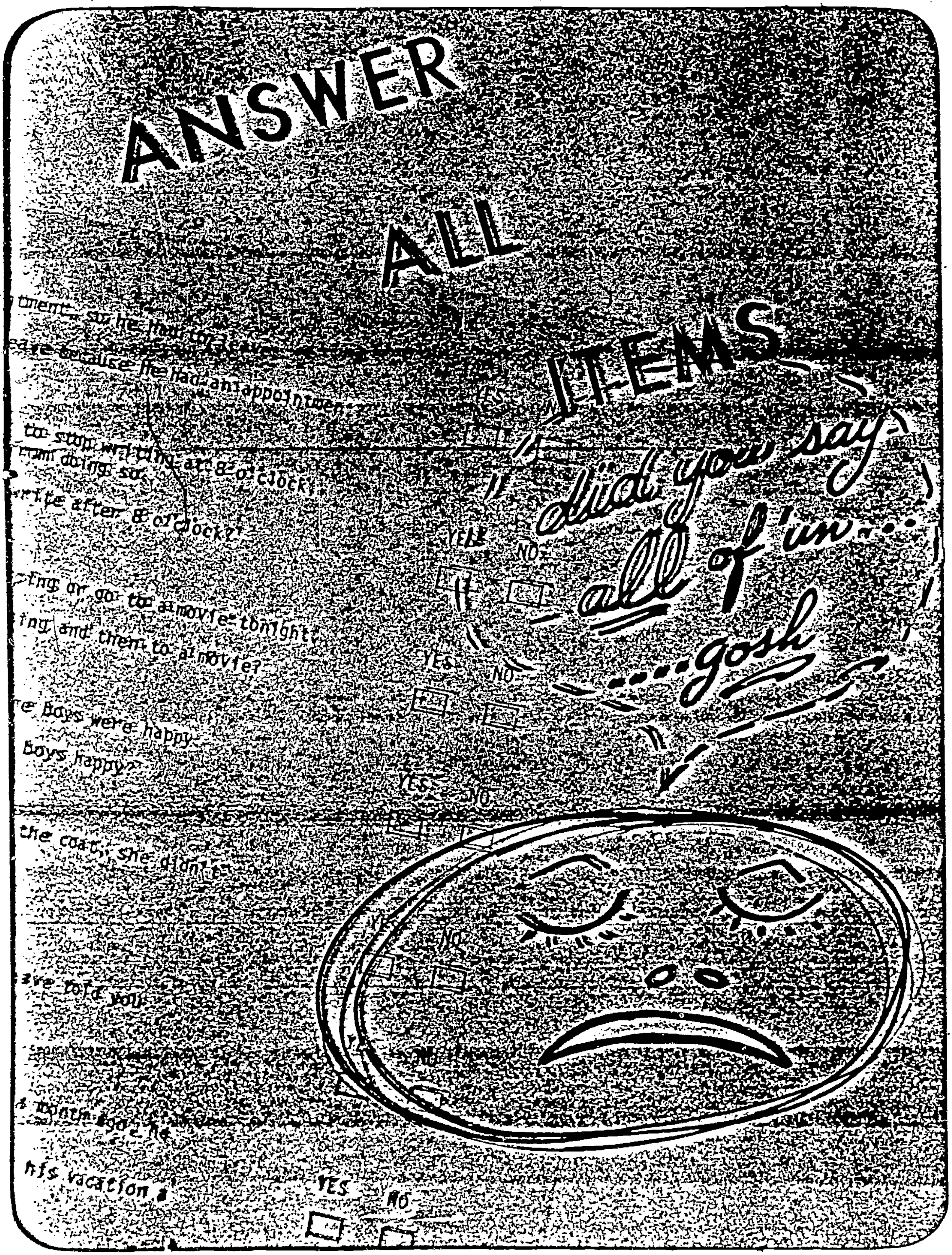
Mary's mother

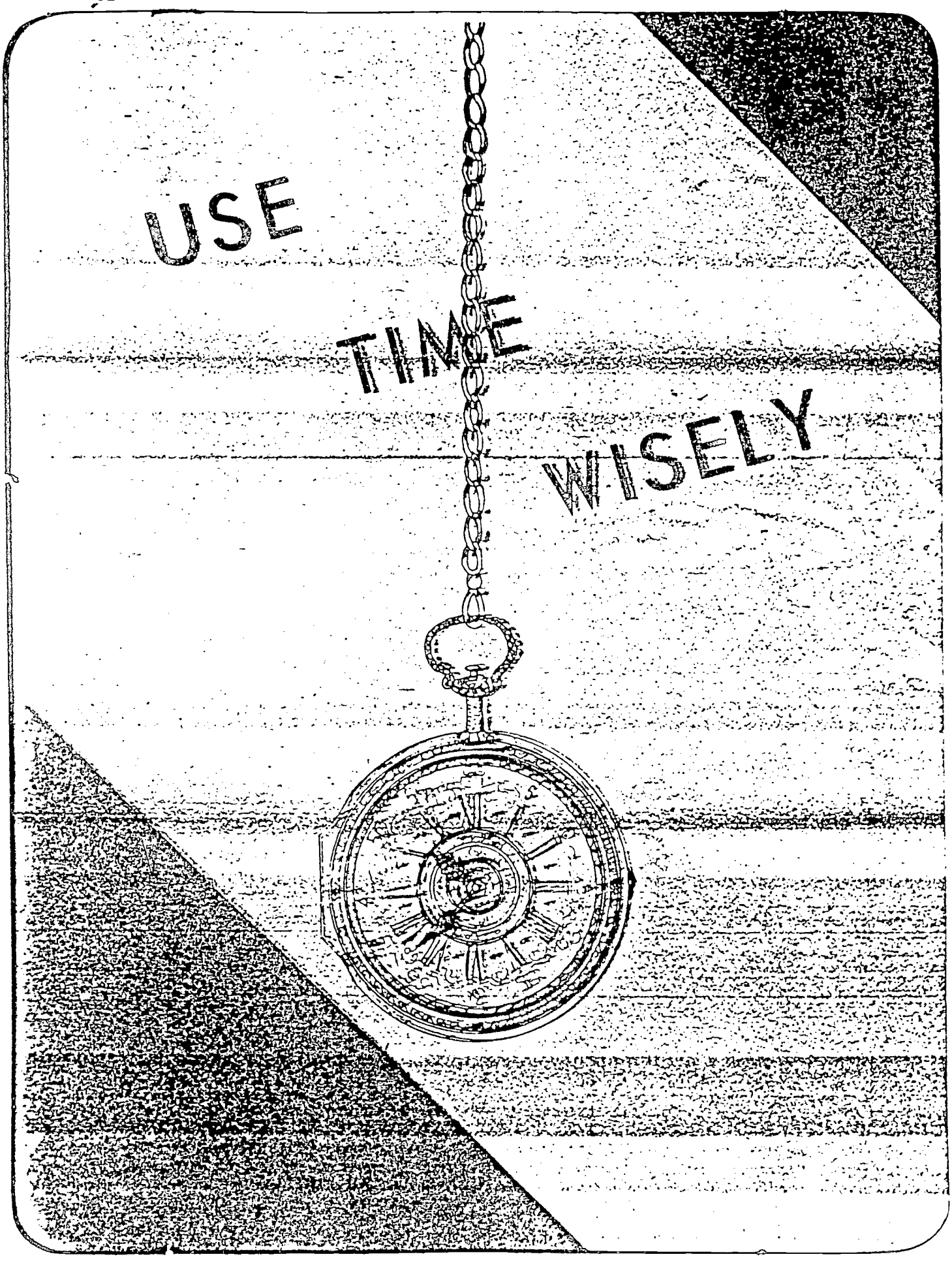
1. is very heavy
2. weighs a lot
*3. is rather thin
4. has too much fat

1, 2, 4 are similar


DON'T
BE A
QUITTER


DON'T LOOK


ANSWER
ALL
ITEMS
did you say
all of 'um...
....gosh


USE
TIME
WISELY


FOLLOW
DIRECTIONS?

APPENDIX 2

The Development and Validation of a Group Test of Testwiseness

Abstract

The validity of the Test of Testwiseness (TOT) was investigated at the third, fifth, and seventh grades following a model for construct validation proposed by Sabers and Whitney. The following properties of the TOT were investigated: (1) internal consistency, (2) stability, (3) sensitivity to instruction, (4) convergent validity, and (5) discriminant validity. Results indicated the TOT was comparatively reliable, relatively stable, sensitive to instruction, and able to identify pupils possessing testwiseness.

The Development and Validation of a Group Test of Testwiseness

Testwiseness (TW) has been recognized since 1951 (Thorndike, 1951) as a significant source of variance in test results. Millman, Bishop, and Ebel (1965) provided the classic definition of TW as "a subject's capacity to utilize the characteristics and formats of the test and/or test-taking situation to receive a high score" (p.707).

The development of the Test of Testwiseness (TOT) followed methods used by Slakter and his associates in their TW scale development (Crehan, Koehler, & Slakter, 1974; Slakter, Koehler, & Hampton, 1970a; Slakter, Koehler, & Hampton, 1970b). In fact, the TOT so closely parallels The Slakter et al scale that it might be considered an elementary level of their test. Four aspects of TW were used in the construction of the TOT. These aspects were chosen from categories described by Millman et al. (1965). These same categories have been selected by other researchers because they were capable of being conveniently measured (Slakter et al., 1970a). Since the TOT was intended to be administered to students in the third, fifth, and seventh grades, many of whom were suspected of low reading ability, the TOT was written with a simple vocabulary. Those aspects of TW which the TOT was intended to measure were described by Millman et al. (1965):

The examinee should be able to

1. Select the option which resembles an aspect of the stem.
2. Eliminate options which are known to be incorrect and to choose from among the remaining options.
3. Eliminate similar options, i.e., options which imply the correctness of each other.
4. Eliminate those options which include specific determiners.

Items were constructed in such a way that the keyed responses could not be arrived at by knowledge of the subject matter. The final form of the scale consisted of 16 items with 4 items designed to measure each of the four categories selected from Millman et al. (1965).

Procedures

The subjects for this study were pupils enrolled in the third (N = 203), fifth (N = 216), and seventh (N = 310) grades of the Tucson Unified School District during the 1980-81 school year. The

ethnic composition of the combined groups consisted of Native Americans (5 percent), Blacks (7 percent), Hispanics (53 percent), and Anglos (35 percent). Males comprised 52 percent of the total group and females 48 percent.

Pupils were pretested with the TOT in the fall of 1980. Following this, volunteer teachers were instructed in TW skills and methods of instruction of TW which they implemented in their classrooms. Students were instructed in principles of TW once every two weeks for a period of four months in 34 experimental classrooms. A total of 9 additional classrooms acted as control groups. Pupils were subsequently posttested with the TOT in the spring of 1981.

Results

Construct validation should include several aspects of reliability and validity. Related to reliability are (1) internal consistency and (2) stability. Aspects of validity which should be examined are (1) sensitivity to instruction, (2) convergent validity, and (3) discriminant validity. These aspects of reliability and validity were investigated following a model proposed by Sabers and Whitney (1976).

Internal Consistency

Internal consistency, an aspect of reliability, measures the extent to which the items measure the same construct. The internal consistency of the TOT was assessed both at the pretest and the posttest five months later using the Kuder-Richardson Formula 20 (KR20). Since the treatment and control groups were different at the posttest time, KR20 coefficients were calculated separately for the experimental and control groups at each grade. The median KR20 coefficient across grades, pretest and posttest was .60. The reliability coefficients for each grade are presented in Table 1 (see next page).

Table 1. Kuder-Richardson Formula 20 Reliabilities and Test-Retest Reliabilities for the 16-item TOT

Grade	N	Kuder-Richardson Formula 20 Pretest	Kuder-Richardson Formula 20 Posttest	Test-Retest Reliabilities
3	203	.398	(E) .612 (C) .479	.49
5	216	.591	(E) .659 (C) .639	.49
7	310	.425	(E) .628 (C) .612	.36

Note: Experimental (E) and Control (C) groups reliability were calculated separately on the posttest.

These results compare favorably with KR20 coefficients reported by Crehan et al. (1974) on the previous 16 item TW scale which was administered in Grades 5-8 in 1968 and again in 1970. The median reliability for the first testing was .40 and for the second testing was .35. In another study, Crehan et al. (1978) reported a median KR20 of .40 on the 16-item TW scale administered to students in grades 5-8.

Stability

Stability is another important aspect of reliability which should be considered in the development of a measure. Stability is estimated with the test-retest method and it reflects the degree to which scores for an individual remain constant over time. Stability coefficients are presented in Table 1. These coefficients suggest a relatively stable measure for a test of this length and type over a five-month period.

Sensitivity to Instruction

Other facets of construct validity investigated in this study were sensitivity to instruction, convergent validity, and discriminant validity. Sensitivity to instruction reflects the extent to which the test can measure the efforts to change the construct in question. To investigate sensitivity to treatment the 34 experimental and 9 control classroom TOT means were used as units of analysis in pre- to posttest comparisons with dependent t-tests. Results are presented in Table 2.

Table 2. Means, Standard Deviations and T-values for Experimental and Control Groups as Measured with the TOT

Group	Number of Classes	Pretest M	Pretest SD	Posttest M	Posttest SD	t
Third						
Experimental	8	9.2	1.0	11.6	1.5	5.91*
Control	3	9.0	.2	8.9	.3	-.41
Fifth						
Experimental	9	9.7	1.1	11.3	1.1	5.05*
Control	2	8.1	.9	8.3	.6	.74
Seventh						
Experimental	17	10.1	.6	12.3	.8	9.72*
Control	4	9.7	.6	10.4	.6	1.54

*$p < .001$

At each grade, experimental groups showed significant pre- to posttest gains while control groups demonstrated no significant gains. These results support the conclusion that the TOT is sensitive to instruction in TW.

Convergent Validity

An examination of convergent validity seeks to determine (1) a correlation between two measures of the same trait, or (2) whether a scale can discriminate between two groups expected to differ on the construct in question (Vaughan & Sabers, 1977). Because no other measure of TW was available to administer to all the students, the contrasted-groups approach was used. After instruction in TW, students in the experimental groups were expected to differ from students in the control groups on TW. Posttest TOT mean scores were compared after they were adjusted for pretest TOT mean differences using analysis of covariance. At each grade level, the experimental groups were significantly higher than the control groups: third grade, $F(1,8) = 11.48$, $p < .01$: fifth grade, $F(1,8) = 5.95$, $p < .05$; and seventh grade, $F(1,18) = 20.50$, $p < .001$. The similarity between this analysis and that used to assess sensitivity to instruction is recognized, but it is the best available measure of convergent validity.

Discriminant Validity

Discriminant validity is another facet which should be investigated. This reflects the degree to which a scale measures a unique construct and not merely reflecting some other construct. Discriminant validity was investigated by correlating the posttest TOT scores with the California Achievement Test (CAT), Form C, Total Reading Test scores. The CAT was administered one month after the posttesting with the TOT. The correlations between the TOT and the CAT appeared to decrease from the third grade, .55 ($p<.001$); fifth grade, .44 ($p<.001$) and seventh grade, .32 ($p<.001$). Comparing these correlations with the TOT's reliability coefficients indicate that the TOT accounts for only a small portion of unique variance at the third grade, but that variance increases from third to the seventh grade. Because the test wise students will use TW to increase their scores on examinations such as the CAT, the above correlations are about as low as should be expected.

Discussion

The present study described the development and validation of the Test of Testwiseness (TOT). This test was constructed in order to test for testwiseness in children in the third, fifth, and seventh grades. The process of item construction, and the investigation of reliability and validity were reported. The results compared favorably with previous research on TW. For a scale of only 16 items, the TOT was found to be comparatively reliable and relatively stable. Moreover, the TOT was able to identify groups possessing TW, and it appeared to be sensitive to instruction.

References

Crehan, K. D.; Gross, L.J.; Koehler, R. A.; and Slakter, M. J. Developmental Aspects of Test-wiseness. Educational Research Quarterly, 1978, 3 (1), 40-44.

Crehan, K. D.; Koehler, R. A.; and Slakter, M. J. Longitudinal Studies of Test-Wiseness. Journal of Educational Measurement, 1974, 11 (2), 209-212.

Millman, J.; Bishop, H.; and Ebel, R. An analysis of test-wiseness. Educational and Psychological Measurement, 1965, 25, 707-726.

Sabers, D. L. and Whitney, D. R. Suggestions for validating scales and attitude inventories (Technical Bulletin No. 19). Iowa City: University of Iowa, Evaluation and Examination Service, 1976 (Mimeographed).

Slakter, M. J.; Koehler, R. A.; and Hampton, S. H. Grade level, sex, and selected aspects of test-wiseness. Journal of Educational Measurement, 1970, 7, 119-122. (a)

Slakter, M. J.; Koehler, R. A.; and Hampton, S. H. Learning test-wiseness by programmed texts. Journal of Educational Measurement, 1970, 7, 247-254. (b)

Thorndike, R. L. Reliability. In E. F. Lindquist (Ed.), Educational Measurement, Washington, D. C.: American Council on Education, 1951.

Vaughan, J. and Sabers, D. Factors in validating affective scales: An applied study. Journal of Reading Behavior, 1977, 9 (3), 253-258.

APPENDIX 3

Test of Testwiseness

APPENDIX 3

STANDARD EDUCATIONAL TEST

Name _____
(Last) (First)

Teacher _____ School _____

Grade _____ Date _____
(Month) (Day) (Year)

Part I

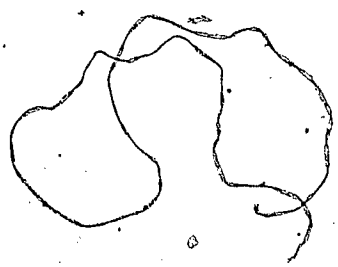
The examiner should read the directions to the students.

Directions:

This part of the test is made up of pairs of words which have either the same or opposite meaning. In front of each pair of words are the letters S for "same" and O for "opposite". Circle the letter S if the pair of words is the same or nearly the same, and circle the letter O for "opposite" if the pair of words is opposite or nearly opposite.

Each correct answer will receive one point. You may leave an answer blank.

S	O	1. glad	happy
S	O	2. sick	ill
S	O	3. plot	frib
S	O	4. strong	weak
S	O	5. hiln	easy
S	O	6. funny	waldy
S	O	7. marnel	mild
S	O	8. flee	fly
S	O	9. sharp	crenic
S	O	10. hot	warm



Stop -- do not go on to the next page

Part II

Examiner should read these directions to the students.

Directions:

Each of the questions or unfinished sentences in this part of the test is followed by four choices. You are to pick the one that best answers the question or finishes the sentence. Circle the letter of your choice. *Answer all problems even if you are not sure of the answer.* Circle your answers in this booklet. Remember, *answer all problems!*

11. "A Birthday Story" is about

A. plants
B. a party
C. wind
D. a rope

12. The King's castle is

E. small
F. tired
G. fast
H. gray

13. What happens when rilts are put in water?

A. heat comes
B. water gets warm
C. water gets blue
D. water gets hot

14. In Pedro City

E. sometimes the sun shines
F. it never rains
G. it only rains in the spring
H. it always rains

15. The flying spider

A. can sing a song
B. can turn green
C. can fly in the air
D. can kill a dog

16. A fly has a

E. frib
F. clock
G. book
H. cake

17. The story "A Long War" is about

A. a small boy
B. an old man
C. a short boy
D. a little boy

18. The game of duki is

E. only played by a king
F. never played
G. sometimes played at home
H. always played

19. Norward means

A. plate
B. backward
C. bake
D. fish

20. Why did Tim smile?

E. Julie was home
F. Joni hurt him
G. He lost his money
H. His TV broke

21. A jimmy worm can

A. jump
B. hop
C. skip
D. eat fish

22. At the party, the bear

E. never eats
F. always drinks
G. sometimes eats fish
H. never wants more

23. The silver-leaf tree

A. has no bugs
B. has large leaves
C. has shiney leaves
D. has smooth branches

24. A policeman carries

E. mace
F. an arrow
G. a spear
H. a sling shot

25. A bear has

A. big legs
B. furry legs
C. large legs
D. fat legs

26. Early Deosophus Man

E. always killed animals
F. never hurt other men
G. always lived in the forest
H. often lived in trees

TEST OF TESTWISENESS (TOT)

KEY

1. Risktaking (Items 1-10)

   Marking an answer for items 3, 5, 6, 7, 9 indicate risktaking.

2. Testwiseness

   A. Stem Option Subscale
      Items: 11(B), 15(C), 19(B), 23(C)

   B. Absurd Option Subscale
      Items: 12(H), 16(E), 20(E), 24(E)

   C. Similar Options Subscale
      Items: 13(C), 17(B), 21(D), 25(B)

   D. Specific Determiners Subscale
      Items: 14(E), 18(G), 22(G), 26(H)

APPENDIX 4

Preliminary Findings:

Report to Participating Teachers and Principals

MEMO

Department of
LEGAL and RESEARCH SERVICES
Division of PLANNING, ANALYSIS and MANAGEMENT

TO: Teachers and Principals Who Participated in NIE Study of Testwiseness 1980-81

FROM: Stephen Powers, Principal Investigator
Darrell Sabers, Consultant, Department of Educational Psychology, University of Arizona

SUBJECT: Preliminary Findings of the National Institute of Education's Testwiseness Study 1980-81

Dr. Sabers and I are happy to send to each participating teacher and principal the Preliminary Findings of the Testwiseness Study 1980-81. In the report there is a teacher number on Tables 2-5 which you can compare with the teacher number on the attached list of participating teachers. This will tell you the results for your class. Since this is a preliminary report, if you find that we inadvertently left the name of a teacher off of the list of participating teachers, please call me at -6138 so that I can make the correction. Five of the teachers who began the TW study chose not to complete the sutyd and so their names were removed from the list of participants.

We were very happy with the results of this study, and thankful for the voluntary participation of the teachers. Without the teacher's assistance, this research could not have been carried out.

Generally the results indicate that (1) teachers implemented very well the testwiseness instruction, (2) students learned the testwiseness principles and (3) the Test of Testwiseness (TOT) did a good job of measuring testwiseness. Teachers administered the TOT as a pretest and a posttest. Actually the first 10 items of the TOT measured Risktaking, and the other 16 items measured Testwiseness. From our analysis, we have concluded the Risktaking measure was not a valid measure, and so our analysis will be directed to the 16 items in the Testwiseness measure.

Some teachers volunteered to participate as Control classes. At first this may seem as a trival participation in the study. Not so! It was actually a crucial part of the study, and a part which the National Institute of Education insisted by incorporated in the Testwiseness Study.

Most of the groups receiving testwiseness training showed significant gains from pre- to posttest while the control groups (which received no testwiseness) did not show significant gains. Therefore, we could conclude the gains in the testwiseness groups were due to testwiseness training and not to maturation, improved reading ability, etc.

To help you understand Tables 2-5:

DOI = Degree Of Implementation: Each teacher either was contacted by phone or contacted through a questionnaire and asked the degree to which they implemented the project.

N = Number Of Students: This is the number of students who participated in the TOT pretest and posttest.

$\overline{X}$ = the mean (or average) of the classroom.

SD = the standard deviation of the scores in the classroom.

r_{xy} = the correlation between the pretest and posttest.

t = this is a value which is obtained as part of the t-test.

p = (A) Probability Value: If this is a number less than or equal to .05, it means that the students gained from pretest to posttest. Any of the following numbers indicated the students gained from pretest to posttest: .000, .048, .025, .001, .016, .003, etc.

(B) It was expected that the p values for students in the Control groups whould not indicate the students gained from pre- to posttest because they were not receiving TW instruction. The following numbers which are larger than .05 indicate students in the Control groups did not make gains: .402, .906, .464, .082.

In summary, in this NIE funded Testwiseness Project 1980-81 students demonstrated that they learned principles of testwiseness. Dr. Sabers and I thank you, the participating teachers, for your voluntary participation and we thank you, the principals at participating schools, for your support.

Sincerely,

Stephen Powers

Stephen Powers and
Darrell Sabers

SP/ch
1/5/82
Enclosure

Copy to Chris Crowder

Teachers and Schools in the NIE Testwiseness Study
1980-81

Teacher	School
Third Grade	
Allen, Doreen	Davidson
Boerner, Shirley	Hollinger
Conway, Ray	Pueblo Gardens
Cotrupe, Clara	Lynn
Darcy, Judith	Pueblo Gardens
Mallams, Elizabeth	Lawrence
Morrison, Carolyn	Safford
Sarver, Marianne	White
Smith, Loise	Pueblo Gardens
Ward, Lopez	Cavett
Westenburg, JoAnne	Drachman
Fifth Grade	
Campas, Gilbert	Drachman
Casey, Donna	Hollinger
Dillard-Davis, Kathy	Lawrence
Ditsworth, Diane	Lynn
Gray, Virginia/Mr. Martinez	Safford
Houston, Sharon	Pueblo Gardens
Kareiva, Anne	Corbett
Regan, Peggy/Mr. Castillo	Van Buskirk
Riggins, Vance	Holladay
Stoler, Nancy	Pueblo Gardens
Thomas, Frances	White
Seventh-Eighth Grade	
Bagwell, Terry	Naylor
Berry, Joann	Wakefield
Born, Diana/Richardson, Gail	Doolen
Cross, Pat	Safford
Freiman, Lela	Naylor
Gordy, Mike	Pistor
Kasulaitis, Rob	Safford
Little, Al	Pistor
Neff-Encinas, Julia	Wakefield
Sizemore, Beverly	Naylor

TUCSON UNIFIED SCHOOL DISTRICT
Department of Legal and Research Services

PRELIMINARY FINDINGS OF THE TESTWISENESS STUDY

1980-81

Prepared by:

Stephen Powers

Stephen Powers, Research Specialist
Legal and Research Services

Approved by:

Christopher Crowder

Christopher Crowder, Assistant
Director, Testing Services

December 1981

MEMO

Department of
LEGAL and RESEARCH SERVICES
Division of PLANNING, ANALYSIS and MANAGEMENT

TO: Christopher Crowder, Assistant Director
Testing Services

FROM: Stephen Powers
Research Specialist

SUBJECT: Preliminary Findings of the Testwiseness Study 1980-81

The National Institute of Education funded a study ($12,191.00) of testwiseness in the Tucson Unified School District for 1980-81. I was the principal investigator and Dr. Darrell Sabers of the Department of Educational Psychology, University of Arizona, was the consultant. One purpose of the study was to implement instruction in testwiseness to volunteering teachers and to observe any gains in students' test-taking skills.

In order to measure testwiseness at Grades 3, 5, 7, and 8, a Test of Testwiseness (TOT) was constructed by Dr. Sabers and myself. The conceptual framework for this test was the most recent research on testwiseness. There were two special problems in the development of this test: (1) it had to be simple enough for third grade students and challenging enough for eighth grade students, and (2) the vocabulary should not be inordinately difficult because many of the students in this study would be minority students coming from elementary and junior high schools of the TUSD where there was low achievement in reading. A further question was, would the test of testwiseness detect changes in students' ability after instruction in testwiseness.

The Test of Testwiseness consisted of five items which measured risktaking in test-taking conditions and 16 items which measured four aspects of testwiseness. The risktaking scale did not appear to be a sensitive or discriminating scale. The testwiseness scale did appear to successfully measure four aspects of testwiseness.

Volunteer teachers were taught testwiseness principles in two one-hour sessions by Dr. Sabers and myself. Teachers then were asked to implement this instruction in their classrooms for a minimum of 15 minutes every two weeks for four months. Students were pretested and posttested with the Test of Testwiseness.

Teachers implemented testwiseness instruction in their own ways in each classroom. For this preliminary analysis, I have analyzed each teacher's classroom of students separately, viewing each group as an independent replication of the treatment. This method of analysis was important because I have promised each teacher that I would report to them the results of the study.

Table 1 summarizes achievement in Testwiseness:

Table 1. Treatment and Control Groups Participating in the Testwiseness Study 1980-81*

	Treatment		Control	
Grade	Classes	Students	Classes	Students
3	8	154	3	49
5	9	186	2	30
7	17	255	4	55
8	13	125	--	--
TOTAL	47	720	9	134

*Only students with pretest and posttest score are included in this analysis.

Of the 47 classrooms receiving testwiseness instruction, 81 percent of the classrooms showed significant gains from pre- to posttest. None of the Control Groups (those groups not receiving testwiseness instruction) showed significant gains from pre- to posttest. Attached to this report are four tables giving detailed analysis of each classroom of students. Only students who were pre- and posttested were included in the analysis.

These results provide evidence for some interesting conclusions:

1. Testwiseness can be taught.
2. Testwiseness can be taught as early as the third grade.
3. A test developed to measure testwiseness will detect changes resulting from testwiseness instruction.
4. Students will not gain in test-taking skills simply by maturation or development of other skills. But specific instruction in testwiseness can raise their testwiseness.

These are only a few preliminary findings. I would like approval to communicate these results to administrators and participating teachers.

Stephen Powers

SP/se
12/2/81
Attachments

Table 2. Dependent t-test Comparing Pretest and Posttest Means of Each TW Participant Teacher's Classes at the Third Grade

Teacher Number	Test	DOI*	N	Pretest $\bar{x}$	Pretest SD	Posttest $\bar{x}$	Posttest SD	r_{xy}	t	P
1	Risk	2	24	4.83	.56	4.96	.20	.06	1.00	.328
	Testwiseness		24	8.79	1.86	11.25	1.73	.37	5.96	.000
2	Risk	2	19	4.84	.37	4.37	1.21	.38	-1.84	.083
	Testwiseness		19	10.37	1.53	12.79	2.42	.61	5.49	.000
6	Risk	2	15	4.47	1.30	5.00	.00	**	1.59	.135
	Testwiseness		15	7.87	2.56	9.33	2.79	.53	2.17	.048
12	Risk	2	23	4.52	1.16	4.91	.417	.28	1.68	.107
	Testwiseness		23	10.43	1.31	12.13	1.14	.45	6.28	.000
14	Risk	1	15	5.00	.00	5.00	.00	**	**	**
	Testwiseness		15	8.33	2.44	12.93	1.58	.32	7.29	.000
33	Risk	1	18	4.89	.32	4.00	1.33	.27	-2.95	.009
	Testwiseness		18	10.11	1.81	12.56	2.33	.58	5.33	.000
38	Risk	1	16	4.50	1.32	5.00	.00	**	1.52	.150
	Testwiseness		16	9.06	2.14	12.75	1.57	.32	6.67	.000
39	Risk	2	24	4.50	1.41	4.96	.20	- .07	1.55	.134
	Testwiseness		24	8.42	2.02	9.42	2.08	.51	2.40	.025
				CONTROL GROUPS						
11	Risk	4	16	4.75	1.00	4.94	.25	1.00	1.00	.333
	Testwiseness		16	9.06	2.64	8.50	2.25	.44	- .86	.402
27	Risk	4	17	4.65	.61	3.88	1.80	- .33	-1.52	.149
	Testwiseness		17	9.12	1.54	9.06	2.63	.65	- .12	.906
35	Risk	4	16	4.94	.25	4.94	.25	- .07	.00	1.000
	Testwiseness		16	8.75	3.02	9.06	2.08	.85	.75	.464

*DOI is Degree of Implementation: Teachers were asked the degree to which they implemented the Testwiseness Project (1 = adequate, 2 = moderate, 3 = slight, 4 = no implementation).

**Cannot be calculated.

Table 3. Dependent t-tests Comparing Pretest and Posttest Means of Each TW Participant Teacher's Classes at the Fifth Grade

Teacher Number	Test	DOI*	N	Pretest $\bar{x}$	Pretest SD	Posttest $\bar{x}$	Posttest SD	r_{xy}	t	P
8	Risk	1	15	4.33	1.76	4.80	.77	.68	1.33	.204
	Testwiseness		15	10.20	2.57	10.73	2.74	.52	.9	.443
9	Risk	3	18	5.00	.00	5.00	.00	**	**	**
	Testwiseness		18	10.00	1.33	10.50	1.15	.31	1.45	.166
15	Risk	1	24	4.79	.72	4.54	1.14	.67	-1.45	.162
	Testwiseness		24	7.50	2.30	9.67	2.35	.41	4.19	.000
16	Risk	1	23	4.17	.94	4.39	1.27	.05	.68	.504
	Testwiseness		23	10.83	1.50	12.70	1.87	.40	4.80	.000
22	Risk	1	23	3.96	1.58	3.35	1.85	.14	-1.30	.208
	Testwiseness		23	8.35	2.95	11.13	2.47	.14	3.74	.001
23	Risk	2	20	4.55	1.28	5.00	.00	**	1.58	.131
	Testwiseness		20	10.15	2.48	10.70	2.15	.22	.85	.409
30	Risk	1	17	4.06	1.34	5.00	.00	**	2.89	.001
	Testwiseness		17	9.94	1.92	11.12	2.17	.75	2.68	.016
32	Risk	1	19	5.00	.00	5.00	.00	**	**	**
	Testwiseness		19	10.21	2.30	13.05	2.25	.28	4.53	.000
37	Risk	2	27	4.89	.32	5.00	.00	**	1.80	.083
	Testwiseness		27	10.52	1.67	12.22	2.76	.33	3.26	.003
				CONTROL GROUP						
20	Risk	4	15	5.00	.00	4.93	.26	**	-1.00	.334
	Testwiseness		15	8.80	3.21	8.73	3.51	.73	- .10	.919
36	Risk	4	15	3.07	2.37	4.80	.56	-.15	2.66	.019
	Testwiseness		15	7.47	3.07	7.93	2.81	.51	.62	.546

*DOI is Degree of Implementation: Teachers were asked the degree to which they implemented the Testwiseness Project (1 = adequate, 2 = moderate, 3 = slight, 4 = no implementation).
**Cannot be calculated.

Table 4. Dependent t-tests Comparing Pretest and Posttest Means of Each TW Participant Teacher's Classes at the Seventh Grade

Teacher Number	Test	Period	DOI*	N	Pretest $\bar{x}$	Pretest SD	Posttest $\bar{x}$	Posttest SD	r_{xy}	t	P
3	Risk	1	2	18	3.39	1.72	3.72	1.84	.74	1.10	.286
	Testwiseness			18	9.56	2.15	13.89	1.18	.42	9.33	.000
5	Risk	3	1	7	3.86	1.46	4.29	1.50	-.13	.51	.629
	Testwiseness			7	10.56	.79	12.29	1.98	-.44	1.87	.111
5	Risk	5	1	9	4.44	.73	4.00	1.50	.57	-1.08	.312
	Testwiseness			9	10.44	2.79	12.78	3.38	.78	3.30	.011
5	Risk	6	1	9	4.89	.33	4.78	.44	-.19	- .55	.594
	Testwiseness			9	8.67	2.24	11.56	2.45	.13	2.80	.023
13	Risk	4	1	17	4.71	.59	5.00	.00	**	2.06	.056
	Testwiseness			17	9.53	2.87	12.65	1.62	.39	4.78	.000
13	Risk	6	1	13	3.31	1.84	4.77	.60	.30	2.99	.011
	Testwiseness			13	10.62	1.56	11.38	1.98	.70	1.95	.075
17	Risk	3	1	4	5.00	.00	5.00	.00	**	**	**
	Testwiseness			4	10.25	2.06	13.25	1.50	.51	3.29	.046
25	Risk	5	2	23	3.43	1.53	4.74	.75	.46	4.59	.000
	Testwiseness			23	11.17	1.82	12.70	2.32	-.30	2.17	.041
25	Risk	6	2	14	4.71	.61	4.86	.53	.34	.81	.435
	Testwiseness			14	93	1.77	11.21	2.19	.40	2.19	.048
28	Risk	1	1	14	4 79	.80	4.43	1.45	-.11	- .77	.455
	Testwiseness			14	9.54	2.56	11.93	2.89	.60	3.47	.004
28	Risk	2	1	21	5.00	.00	4.33	1.68	**	1.81	.085
	Testwiseness			21	9.95	3.11	11.05	3.32	.73	2.14	.045

Table 4. (contd.)

Teacher Number	Test	Period	DOI*	N	Pretest $\bar{x}$	Pretest SD	Posttest $\bar{x}$	Posttest SD	r_{xy}	t	P
31	Risk	1	1	23	5.00	.00	4.96	.21	**	-1.00	.328
	Testwiseness			23	9.39	1.50	12.83	2.06	.38	8.07	.000
31	Risk	2	1	16	4.37	.96	4.81	.75	.10	1.52	.150
	Testwiseness			16	10.12	2.06	12.75	2.16	.50	4.94	.000
31	Risk	3	1	13	4.54	1.13	4.62	.96	-.18	.17	.866
	Testwiseness			13	10.08	2.60	12.69	2.90	.34	2.97	.012
31	Risk	4	1	22	4.64	.790	4.91	.29	.26	1.67	.110
	Testwiseness			22	10.27	1.39	11.91	2.65	.20	2.81	.011
31	Risk	5	1	21	4.76	.89	5.00	.00	**	1.23	.234
	Testwiseness			21	9.90	1.55	12.71	2.19	.17	5.23	.000
34	Risk	1	3	12	4.58	1.00	5.00	.00	**	1.45	.175
	Testwiseness			12	10.92	1.31	12.08	1.56	.36	2.46	.032
					CONTROL GROUPS						
5	Risk	4	4	15	4.93	.26	3.00	2.27	-.12	-3.24	.001
	Testwiseness			15	10.60	1.80	10.27	2.31	.63	- .70	.494
19	Risk	2	4	11	5.00	.00	5.00	.00	**	**	**
	Testwiseness			11	9.36	2.29	10.64	2.94	.30	1.35	.210
19	Risk	6	4	7	5.00	.00	5.00	.00	**	**	**
	Testwiseness			7	9.57	1.40	11.00	1.63	.29	2.09	.082
25	Risk	4	4	22	4.54	1.01	4.82	.50	.58	1.55	.137
	Testwiseness			22	9.36	2.10	9.59	2.82	.43	.39	.700

*DOI is Degree of Implementation: Teachers were asked the degree to which they implemented the Testwiseness Project (1 = adequate, 2 = moderate, 3 = slight, 4 = no implementation).
**Cannot be calculated.

Table 5. Dependent t-tests Comparing Pretest and Posttest Means of Each TW Participant Teacher's Classes at the Eighth Grade

Teacher Number	Test	Period	DOI*	N	Pretest $\bar{x}$	Pretest SD	Posttest $\bar{x}$	Posttest SD	r_{xy}	t	p
3	Risk	3	1	15	3.87	1.77	4.93	.26	-.18	2.26	.041
	Testwiseness			15	11.27	1.79	13.60	1.96	.28	4.00	.001
3	Risk	4	1	11	5.00	.00	5.00	.00	**	**	**
	Testwiseness			11	10.55	1.81	14.27	1.27	-.07	5.41	.000
3	Risk	6	1	14	4.86	.53	4.86	.53	-.08	.00	1.00
	Testwiseness			14	10.29	1.33	13.71	1.54	-.26	5.63	.000
13	Risk	1	1	13	4.54	1.39	4.92	.28	.98	1.24	.240
	Testwiseness			13	9.69	2.46	10.54	2.26	.30	1.09	.296
13	Risk	2	1	10	5.00	.00	4.80	.42	**	-1.50	.168
	Testwiseness			10	9.30	2.31	12.00	1.70	-.31	2.61	.028
13	Risk	5	1	12	4.58	1.16	4.75	.62	-.16	.41	.689
	Testwiseness			12	10.17	1.33	12.67	1.97	-.32	3.19	.009
17	Risk	6	1	5	4.60	.55	5.00	.00	**	1.63	.178
	Testwiseness			5	8.40	1.52	15.00	1.41	.82	16.50	.000
24	Risk	1	1	6	5.00	.00	5.00	.00	**	**	**
	Testwiseness			6	10.00	1.79	13.33	1.37	-.41	3.07	.028
24	Risk	2	1	5	5.00	.00	5.00	.00	**	**	**
	Testwiseness			5	10.80	1.79	11.40	5.81	-.69	.19	.861

Table 5. (contd.)

Teacher Number	Test	Period	DOI*	N	Pretest $\bar{x}$	Pretest SD	Posttest $\bar{x}$	Posttest SD	r_{xy}	t	P
24	Risk	5	1	7	5.00	.00	4.86	.38	**	-1.00	.356
	Testwiseness			7	9.71	1.80	12.57	1.72	.65	5.16	.002
24	Risk	6	1	7	4.86	.38	5.00	.00	**	1.00	.356
	Testwiseness			7	10.29	1.80	12.00	1.83	.41	2.30	.061
28	Risk	3	1	16	4.69	1.01	4.94	.25	-.08	.94	.362
	Testwiseness			16	10.56	2.39	13.56	2.25	.44	4.90	.000
28	Risk	4	1	4	4.75	.50	5.00	.00	**	1.00	.391
	Testwiseness			4	11.25	1.26	13.00	.82	-.32	2.05	.133

*DOI is Degree of Implementation: Teachers were asked the degree to which they implemented the Testwiseness Project (1 = adequate, 2 = moderate, 3 = slight, 4 = no implementation).

**Cannot be calculated.

NOTE: There were no control groups for eighth grade students.

APPENDIX 5

Repeated Measures

Analysis of Variance on Reading Comprehension

Scores

Table 1. Two Factor Repeated Measures ANOVA on Reading Comprehension with Third Grade Students*

Source of Variation	df	MS	F	p
Between subjects				
Ethnicity (E)	3	83.59	3.97	.010
Subj w. groups	103	21.04		
Within subjects				
Reading (R)	1	35.09	8.19	.005
RE	3	5.35	1.25	.296
Rx sub w. groups	103	4.28		

*Although a three factor Repeated Measures ANOVA was planned for the third grade, the high SES Native American cell did not have any cases. Therefore, the three factor analysis could not be completed.

Table 2. Three Factor Repeated Measures ANOVA on Reading Comprehension with Fifth Grade Students

Source of Variation	df	MS	F	p
Between subjects				
Ethnicity (E)	3	924.48	6.69	.000
SES (S)	1	269.50	1.95	.164
ES	3	18.63	.13	.939
Subj w. groups	171	138.20		
Within subjects				
Reading (R)	1	.76	.09	.763
RE	3	21.32	2.55	.057
RS	1	4.33	.52	.473
RES	3	11.31	1.35	.259
RX sub w. groups	171	8.36		

Table 3. Three Factor Repeated Measures ANOVA on Reading Comprehension with Seventh Grade Students

Source of Variation	df	MS	F	p
Between subjects				
Ethnicity (E)	3	800.58	6.74	.000
SES (S)	1	.40	.00	.954
ES	3	136.43	1.15	.331
Subj w. groups	191	118.80		
Within subjects				
Reading (R)	1	5.57	.46	.496
RE	3	17.10	1.43	.237
RS	1	.10	.01	.926
RES	3	9.59	.80	.496
RX subj w. groups	191	11.99		

Table 4. Descriptive Statistics of Third Grade Students by Ethnicity and Time (Pretest/Posttest) on Reading Comprehension

Ethnicity	N	Pretest Mean	SD	Posttest Mean	SD	Total Mean
Native American	5	31.77	3.62	34.12	2.28	32.95
Black	15	34.59	3.93	35.93	3.53	35.26
Hispanic	51	35.79	3.27	36.00	3.53	35.90
Anglo	36	37.03	4.16	37.83	3.30	37.43
TOTAL	107	35.85		36.52		36.19

Table 5. Descriptive Statistics of Fifth Grade Students by Ethnicity and Time (Pretest/Posttest) on Reading Comprehension.

Ethnicity	N	Pretest Mean	SD	Posttest Mean	SD	Total Mean
Native American	3	43.76	5.04	41.44	3.67	42.60
Black	6	34.80	3.25	36.54	4.05	35.67
Hispanic	98	43.61	7.74	44.61	8.50	44.11
Anglo	72	49.98	9.97	50.15	9.15	50.06
TOTAL	179	45.88		46.52		46.20

Table 6. Descriptive Statistics of Seventh Grade Students by Ethnicity and Time (Pretest/Posttest) on Reading Comprehension.

Ethnicity	N	Pretest Mean	SD	Posttest Mean	SD	Total Mean
Native American	7	43.58	8.79	41.47	6.65	42.53
Black	14	43.72	7.92	43.71	7.31	43.72
Hispanic	107	43.73	7.26	46.02	7.74	44.88
Anglo	71	50.73	9.31	51.75	8.81	51.24
TOTAL	199	46.23		47.74		46.98

REFERENCES

Arizona Daily Star, 100 School Districts Join Test-Help Plan, August 14, 1981, Section B, Page 2.

Austin, Linda. Teachers pressured to Teach Test-taking Skills. Dallas Times Herald, January 26, 1981. Metro Page.

Bajtelsmit, J. W. Development and Validation of an Adult Measure of Secondary Cue-Using Strategies on Objective Examinations: The Test of Obscure Knowledge (TOOK). Paper presented at the annual meeting of the Northeastern Educational Research Association, Ellenville, New York, October 1975a.

Callenbach, C. The Effects of Instruction and Practice in Content-Independent Test-Taking Techniques Upon the Standardized Reading Test Scores of Selected Second Grade Students. Journal of Educational Measurement, 1973, 10, pp. 25-30.

Crehan, K. D., Koehler, R. A., and Slakter, M. J. Longitudinal Studies of Test-wiseness. Journal of Educational Measurement, 1974, 11, pp. 209-212.

Gaines, W. G., and Jongsma, E. A. The Effect of Training in Test-Taking Skills on the Achievement Scores of Fifth Grade Pupils. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, Illinois, April 1974.

Gibb, B. G. Test-wiseness as Secondary Cue Response. (Doctoral dissertation, Stanford University), Ann Arbor, Michigan: University Microfilms, 1964, No. 64-7643.

Gifford, Charles S. and Fluitt, John L. How to make your students test-wise. The American School Board Journal, 167 (10), October 1980, 29.

Langer, G., Wark, D., and Johnson, S. Test-wiseness in Objective Tests. In PL Nacke (Ed.), Diversity in Mature Reading: Theory and Research, Vol. I, 22nd Yearbook of the National Reading Conference. National Reading Conference, Milwaukee, Wisconsin, 1973.

McPhail, Irving P. Why teach Testwiseness? Journal of Reading 25(1), October 1981, 32-38.

Millman, J. Test-wiseness in Taking Objective Achievement and Aptitude Examinations. Final Report, 1966, College Entrance Examination Board.

Millman, J., Bishop, C. H., and Ebel, R. An Analysis of Test-wiseness. Educational and Psychological Measurement, 1965, 25, pp. 707-726.

Moore, J. C., Schutz, R. E., and Baker R. L. The Application of a Self-Instructional Technique to Develop a Test-Taking Strategy. American Education Research Journal, 1966, 3, pp. 13-17.

Oakland, T. The Effects of Test-wiseness Materials on Standardized Test Performance of Preschool Disadvantaged Children. Journal of School Psychology, 1972, 10, pp. 355-360.

Sabers, D. Test-Taking Skills, Arizona Center for Educational Research and Development, College of Education, University of Arizona, Tucson, Arizona, 1975.

Sarnacki, R. E. An Examination of Test-wiseness in the Cognitive Test Domain. Review of Educational Research, 49(2), p. 255.

Slakter, M. J., Keohler, R. A., and Hampton, S. H. Grade Level, Sex, and Selected Aspects of Test-wiseness. Journal of Educational Measurement, 1970a, 7, pp. 119-122.

Slakter, M. J., Keohler, R. A., and Hampton, S. H. Learning Test-wiseness by Programmed Texts. Journal of Educational Measurement, 1970b, 7, pp. 247-254.

Thorndike, R. L. Reliability. In E. F. Lindquist (Ed.), Educational Measurement, Washington, D. C.: American Council on Education, 1951.

Woodley, K. K. Test-wiseness Program Development and Evaluation. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, Louisiana, 1973.