

AN EVALUATION OF THE CONCORDIA PBK INVENTORY

by

WAYNE HERMAN WENTZEL

B.S., Concordia Teachers College
Seward, Nebraska, 1965

A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

College of Education

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1969

Approved by:


Major Professor

LD
2668
T4
1969
W45
C.2

ACKNOWLEDGEMENTS

I wish to acknowledge the efforts of Dr. Gerald Hanna, Kansas State University, who spent many hours helping me determine how to use and interpret the materials at my disposal. I also wish to thank my good wife for her help and encouragement in gathering and assembling materials for this paper.

TABLE OF CONTENTS

CHAPTER	PAGE
I. INTRODUCTION	1
Purpose of the study	1
Limitations of the study	4
Preview of the study	4
Background of the test	5
II. SAMPLE AND PROCEDURE	8
Selection of the sample	8
Procedure	8
III. FINDINGS AND DISCUSSION	13
IV. SUMMARY	25
BIBLIOGRAPHY	28

LIST OF TABLES

TABLE	PAGE
I. Split-Half Coefficients of Reliability for the Total Score of Both Forms of the Test at Each Grade Level	14
II. Split-Half Coefficients of Reliability for Each Section of Both Forms of the Test at Each Grade Level	15
III. Parallel Form Reliability at Each Grade Comparing Maehr Total Score and Simple Total Score	17
IV. Correlation Coefficients at Each Grade Level Comparing the Maehr Scoring Method with the Simple Total	17
V. Item Discrimination Index, Forms X and Y, Grade Five	19
VI. Item Discrimination Index, Forms X and Y, Grade Six	20
VII. Item Discrimination Index, Forms X and Y, Grade Seven	21
VIII. Item Discrimination Index, Forms X and Y, Grade Eight	22
IX. Results of the Fakability Study, Form X, Grades Five through Eight	23

CHAPTER I

INTRODUCTION

Purpose of the study. In the fall of 1966, the Concordia PBK Inventory, written by Dr. Martin J. Maehr of Concordia Teachers College, Seward, Nebraska, was administered in approximately seventeen Lutheran schools in the state of Kansas. The purpose of this administration of the Inventory was to introduce the test to various schools in order to aid in the determination of its usefulness as an instrument for measuring religious knowledge and attitudes.

It was noted by several men related to Lutheran education, including the author of the test, that little or nothing had been done with the results of the administration of the test. The initial purpose of this study was to use these results toward the establishment of norms which could be used in the Lutheran schools of Kansas as a basis for comparison.

The review of the test manual was the first part of the study undertaken. Standards for Educational and Psychological Tests and Manuals¹ was used as the basis for this part of the study. As these standards were compared with

¹American Psychological Association, Standards for Educational and Psychological Tests and Manuals (Washington: American Psychological Association, 1966).

the material recorded in the PBK Inventory manual, the following items were noted:

1. The manual did not include what universe of students was represented as a sample. Consequently, the user of the test would not know from what area or group of children the sample was taken.²

2. No evidence was given as to the extent to which the scores were susceptible to an attempt on the part of the student to present a favorable picture of himself.³

3. There was no evidence reported of the finding of the mean and standard deviation for the sample from which the coefficients of reliability given in the test manual were obtained. Knowledge of the mean and standard deviation are necessary to facilitate proper interpretation of the reliability coefficients.⁴

4. The manual did not report whether the reliability analysis was based on children in a single grade or over a multi-grade sample, nor was the method of selection of the sample given.⁵

5. There was no evidence reported of internal consistency. Test-retest reliability and parallel form reliability were reported, however internal consistency should

²American Psychological Association, op. cit., p. 15.

³Ibid., p. 24.

⁴Ibid., p. 28.

⁵Ibid., p. 28.

also have been indicated.⁶

6. The directions printed in the manual and on the pupil booklets were not complete enough for the pupil to understand the intentions of the author of the test. The pupils were not instructed through reading the directions of the importance of honesty in answering Practice and Belief items.⁷

7. Since the test results could have been recorded either in the test booklet or on separate answer sheets, no data were reported in the manual to show the extent to which these methods were interchangeable.⁸

8. The manual did not establish a rationale for the unusual scoring system used for this test. Each form of the test had twenty-six topical areas of three items each. A total test score was derived by adding three points if all items in a section were correct, two points if two items were correct, but no points if only one item was correct. No explanation was offered in the manual as to why such single responses were not included in the score.⁹

9. No indication was given in the manual as to whether the primary purpose of the test was to compare individuals with their local group or with a larger reference group. If

⁶Ibid., p. 30

⁸Ibid., p. 33.

⁷Ibid., p. 32.

⁹Ibid., p. 33.

the latter was intended, norms for such a reference were lacking.¹⁰ Norms which did exist were provided to categorize the results of the test as poor, low average, average, high average, good or excellent. The sample used to establish these norms was described as a "representative Lutheran elementary school population of grades 6-8,"¹¹ but the size and locality distribution of the sample were not reported.

Due to the several areas in which the manual failed to report information considered in Standards for Educational and Psychological Tests and Manuals to be essential or desirable, the purpose of this study is to provide such information.

Limitations of the study. The study is limited to data provided by Lutheran schools in the state of Kansas which participated in the testing program. Since all scoring of the test was done by hand, a margin of error must be taken into account.

Preview of the study. The study will concern itself with a discussion of the test, the selection of the sample used, the procedure used in collecting and evaluating the

¹⁰Ibid., p. 34.

¹¹Martin J. Maehr, Manual to Accompany Concordia PBK Inventory (St. Louis: Concordia Publishing House, 1965), p. 7.

acquired data, a report of the findings of the study, and a final chapter dealing with a summary of the developments of the previous chapters and a restatement of important findings in the study.

Background of the test. The Concordia PBK Inventory is a test designed to assess the outcomes of religious education. The author of the test stated its purpose as follows:

The Concordia Practice, Belief, Knowledge (PBK) Inventory was prepared to discover relationships existing between Bible information which the students at the upper elementary age possess on the one hand and the kinds of practices and beliefs they express on the other.¹²

The test contains three separate sections, with twenty-six items in each section. The first, section P, "is designed to sample the pupil's reaction to contemporary situations that parallel stories and incidents of the Old and the New Testament."¹³ The second, section B, "attempts to elicit the student's response on what he considers to be correct practice in relation to the comparable conduct item."¹⁴ The third, section K, "samples the pupil's acquaintance with the representative Bible stories or incidents from which the respective practice and belief inventory items were drawn."¹⁵

¹²Ibid., p. 3.

¹⁴Ibid., p. 3.

¹³Ibid.

¹⁵Ibid.

Each section is made up of twenty-six multiple choice items, with four possible answers provided for each item. Thus, the entire test consists of seventy-eight multiple choice items. The test is constructed in such a way that each item of section P is based on the same Biblical or moral teaching as the corresponding items of B and K. The following items were taken from Form Y of the test to illustrate how three corresponding items of each topical triad relate to each other. They each deal with the subject of forgiving those who offend.

P Part I

2. When someone hurts my feelings and then asks for forgiveness:
- 1. I wait until he pleads for forgiveness.
 - 2. I forgive him.
 - 3. I refuse until he proves that he is sorry.
 - 4. I remind him of how bad he was.

B Part II

2. Mike isn't a bad boy, but he has the habit of saying mean things to his friends. I believe Mike's friends should:
- 1. Make him pay.
 - 2. Forgive him whenever he asks.
 - 3. Not speak to him again.
 - 4. Warn other children against Mike.

K Part III

2. In Jesus' parable about the unmerciful servant, that certain servant was immediately:
- 1. Thrown into prison.
 - 2. Forgiven.
 - 3. Warned about making debts.
 - 4. Sent away.

The results of the test are placed on a pupil analysis sheet which is divided into twenty-six columns, each of which is in turn subdivided into three additional columns. Each of the twenty-six columns is for scoring a different Biblical or moral teaching, such as is illustrated by the items above. The three subdivisions are for scoring P, B, and K of each topical division. The total number of the twenty-six divisions under which all three, B, P, and K, are marked as correct is multiplied by three. The number of divisions under which any combination of two, such as PB, BK, or PK, is marked as correct is multiplied by two. No allowance is made in the prescribed scoring method for a single correct response under a given division. The total score for the test can be illustrated as follows:

$$T = (PBK \times 3) + (PB \times 2) + (PK \times 2) + (BK \times 2).$$

CHAPTER II

SAMPLE AND PROCEDURE

Selection of the sample. Six hundred and eighty-nine pupils from seventeen Lutheran schools in the state of Kansas were used as the sample for this study. Although a larger number of students participated in the testing program from which the sample was taken, only those pupils were used whose scores were available on both forms of the test. This number represented 95 per cent of the total number who took the test. The sample represents students from grades five through eight with the following breakdown by grade:

grade 5--164 pupils
grade 6--186 pupils
grade 7--175 pupils
grade 8--164 pupils.

Procedure. The results of the 1966 administration of the PBK Inventory were obtained through the office of the Lutheran Church-Missouri Synod, Kansas District, in Topeka. The results, including the name, school and grade level of each child in addition to his correct or incorrect response on each of the seventy-eight items on both forms of the test, were placed on IBM cards. Each student was given an identification number and a code number for his grade, sex, and the size of the community at which his school was located. Community size was divided into two groups, large and small,

with a population of ten thousand used as the separation point between the two. The response by each child to each item was coded on the basis of whether the response was correct or incorrect. The computer was programed to provide inter-correlations with all of the variables on the IBM cards.

From the material which was coded on the IBM cards, the following information was computed:

1. The coefficient of reliability using the split-half method was found for each of the three sections of the test. These coefficients were found for both forms of the test at each grade level.

2. The parallel form coefficient of reliability was found for each section (P, B, and K) at each grade level at which the test was administered.

3. Because of the unusual scoring method prescribed by the author of the test, a method which did not allow for credit if only one response out of the three sections was correctly answered, all of the papers were rescored to find a simple number of correct answers. This simple score was correlated with the score achieved by the method prescribed in the test manual.

4. The point bi-serial correlation coefficient of each item with the score of the section to which it belonged was computed. These item discrimination indices were

computed separately for each grade level.

On most tests which deal with the measurement of a pupil's attitude, it is necessary to determine the extent to which the test is fakable. In other words, it must be determined if the test will allow the pupil to present a favorable picture of himself even though such a picture might be false. In order to determine the fakability of the test, it was administered to thirty-two children of grades five through eight in attendance at Immanuel Lutheran School in Junction City, Kansas. Form X of the test was administered twice to each of these children. They were divided into two groups, with grades five and six forming one group, and grades seven and eight forming the other. A rotated design was employed in this administration. Such a design calls for the first set of directions to be read to one group of pupils while the second set of directions is being read to the other group. This order is then reversed for the second administration. This rotation of the test directions is intended to equalize practice effect which results from taking a test twice within a time span of only six days.

The two sets of directions for the test written for this study were designed to compare the difference between pupils' scores when on one hand they were told they would be graded on the test and should strive for a good grade, and on the other hand when they were told to answer only as

they believed or practiced.

The directions pertaining to the pupils which were found in the test manual were read to each group prior to both administrations. They read as follows:¹⁶

4. Say: "You are to decide whether you are going to mark choice 1, 2, 3, or 4. Mark an (X) before the choice you select. Look at the sample."

Sample:

1. When I am sleepy I want to:

- 1. Play.
- (X) 2. Rest.
- 3. Eat.
- 4. Run.

"If No. 2 in this sample is chosen, then you mark an (X), like this. Mark only ONE answer of the four given for EACH of the 26 items in each section of the inventory."

5. Say next: "Answer all the questions in part one. Wait until you are told to go on to part II or to part III. Are you ready? Go!"

In addition to these directions, the following were read aloud to the pupils:

1. Listen very carefully as I read the following directions to you. The directions will be read only once. You will not be given a chance to ask questions for any reason. You are not being graded on this test. It is important that you mark the answer which you believe or feel rather than that which you know to be the most desirable

¹⁶Maehr, op. cit., p. 4.

response.

2. Listen very carefully as I read the following directions to you. The directions will be read only once. You will not be given a chance to ask questions for any reason. I am giving you this test to determine your knowledge of how a good Christian would believe or practice. You will be graded on your work so be sure that you mark your answers carefully in order to get the highest score you possibly can.

Prior to the second administration of the test, the pupils to whom the first set of directions was read were told that it had been decided to re-administer the test in order that they might be graded on it. The pupils to whom the second set of directions was read were told that the first administration of the test had been faulty and that they would not be held responsible for the second administration.

CHAPTER III

FINDINGS AND DISCUSSION

The Practice and Belief sections of the test deal with the affective domain, being basically personality inventories. They are designed to reflect the way individuals act and think toward and about people, objects and situations they encounter as a result of their previous experiences. The split-half reliability coefficients shown in Table I compare favorably with standardized achievement tests. Since the reliabilities of achievement batteries would normally be higher than those of personality inventories, the reliability coefficients shown in Table I appear to be quite good. Notably higher values tend to exist as the grade level at which the test was given gets higher.

The reliability coefficients reported in Table I were computed by using alternate triads. Each triad consisted of the P, B, and K items under a given number. Of the twenty-six topical areas in the test, thirteen odd and thirteen even were used, each consisting of a triad P, B, and K.

Table II indicates that the split-half coefficients of reliability, when found for each of the three sections of the test, were somewhat lower. Most of the coefficients again indicated higher reliability as the grade level became

higher. It is interesting to note that the coefficients of reliability were found to be lowest for the knowledge section, which is actually an achievement test. This seems to be in contrast to a statement made by Ahman and Glock that in relation to achievement tests, "values for personality inventories very considerably but frequently are lower."¹⁷

TABLE I

SPLIT-HALF COEFFICIENTS OF RELIABILITY FOR THE TOTAL SCORE OF BOTH FORMS OF THE TEST AT EACH GRADE LEVEL

		Grade 5 N=164	Grade 6 N=186	Grade 7 N=175	Grade 8 N=164
Form X	Mean	51.4	54.5	58.7	58.9
	S.D.	15.7	13.7	15.5	17.3
Form Y	Mean	50.0	54.8	58.8	60.0
	S.D.	16.0	15.7	15.1	17.2
Total Score Reliability	X	.91	.92	.95	.96
	Y	.93	.94	.92	.96

¹⁷J. Stanley Ahmann and Marvin D. Glock, Evaluating Pupil Growth (Boston: Allyn and Bacon, Inc., 1967), p. 328.

TABLE II
 SPLIT-HALF COEFFICIENTS OF RELIABILITY FOR
 EACH SECTION OF BOTH FORMS OF THE
 TEST AT EACH GRADE LEVEL

		Grade 5 N=164	Grade 6 N=186	Grade 7 N=175	Grade 8 N=164
Form X P	r	.90	.88	.90	.83
	M	21.5	22.3	22.4	21.9
	S.D.	4.7	4.2	4.9	5.0
Form Y P	r	.82	.84	.79	.78
	M	19.9	21.0	21.5	21.2
	S.D.	4.2	4.4	4.4	4.9
Form X B	r	.87	.85	.94	.95
	M	20.5	21.7	22.6	22.1
	S.D.	4.9	4.1	4.7	5.7
Form Y B	r	.88	.91	.93	.95
	M	19.6	21.1	22.0	22.1
	S.D.	5.3	4.8	4.6	5.5
Form X K	r	.77	.78	.87	.91
	M	13.3	14.1	16.2	17.3
	S.D.	4.7	4.5	5.7	6.1
Form Y K	r	.80	.79	.87	.91
	M	14.1	15.2	17.1	18.2
	S.D.	4.8	4.9	5.7	6.1

Forms X and Y of the test are not actually parallel forms, since one tests Old Testament and the other New Testament. Also, the corresponding items on the two forms do not deal with the same topic, such as items numbered one dealing with selfishness, items numbered two dealing with loyalty, and so on. Therefore, the parallel form coefficient of reliability is reported with some reservation. The results in Table III indicate that the two forms correlate relatively high. The coefficients are reported for both the Maehr method of scoring and the simple total score. When comparing the results reported in Table III with those of Table I, the parallel form reliability is somewhat lower. This would seem to indicate that even though the internal consistency of the test is high, the stability is more limited.

Table III also indicates a high similarity in parallel form reliability between scores based on Maehr's method of scoring and those based on a simple total. In fact, those found using the simple total are higher. This would indicate that since the author of the test does not give his readers any reason or justification for using his method of scoring, there seems to be no apparent advantage to its use. Table IV verifies this statement by pointing out the extremely high correlation between Maehr's method of scoring and a simple total.

TABLE III

PARALLEL FORM RELIABILITY AT EACH GRADE COMPARING
MAEHR TOTAL SCORE AND SIMPLE TOTAL SCORE

		Grade 5 N=164	Grade 6 N=186	Grade 7 N=175	Grade 8 N=164
r	X Y	.80	.81	.74	.85
Maehr Total	M	51.4	54.5	58.7	58.9
	X Y	50.0	54.8	58.9	60.0
S.D.	X	15.7	13.7	15.5	17.3
	X Y	16.0	15.6	15.1	17.2
r	X Y	.81	.81	.78	.86
Simple Total	M	55.4	58.0	61.2	61.3
	X Y	53.6	57.3	60.6	61.4
S.D.	X	12.3	10.5	12.9	14.8
	X Y	12.5	12.2	12.8	14.6

TABLE IV

CORRELATION COEFFICIENTS AT EACH GRADE LEVEL COMPARING
THE MAEHR SCORING METHOD WITH THE SIMPLE TOTAL

	Grade 5 N=164	Grade 6 N=186	Grade 7 N=175	Grade 8 N=164
Form X	.99	.99	.99	.997
Form Y	.99	.995	.99	.998

An item discrimination index is intended to show the relationship between the responses by members of the sample group on each item and the total score of the section of the test. Item one from section P, form X, was correlated with the total of section P, form X, and so on. The item discrimination indices which were done at each grade level indicated no specific patterns as far as progression by grade level is concerned, however indices at grades seven and eight tend to be higher. Ahmann and Glock state that "values less than 0.20 indicate that the discrimination power of the test item is questionable."¹⁸ Although there are scattered examples of items with a point bi-serial correlation of less than .20, only items numbered twenty-five and twenty-six show any consistency below that point. These questions deal with cursing, obedience and respect to parents, respect for the teacher, and the sin of offense. Items under triad twenty-five in several instances reveal a negative correlation. This means that in these instances, pupils who scored higher on the total test missed these questions more frequently than those scoring lower. Many of the items, however, have quite high correlations of .70 and above. This indicates that most items were responded to with relatively high consistency and were discriminating quite adequately.

¹⁸Ahmann and Glock, op. cit., p. 189.

TABLE V
 ITEM DISCRIMINATION INDEX, FORMS X AND Y,
 GRADE FIVE (N=164)

Item	P X	B X	K X	P Y	B Y	K Y
1	.36	.45	.33	.30	.53	.38
2	.47	.37	.27	.42	.47	.14
3	.47	.45	.18	.50	.59	.50
4	.45	.44	.54	.27	.38	.14
5	.42	.65	.22	.44	.51	.50
6	.51	.69	.47	.28	.42	.23
7	.49	.77	.40	.33	.43	.41
8	.25	.45	.45	.58	.76	.51
9	.65	.70	.48	.36	.53	.56
10	.50	.48	.50	.51	.57	.42
11	.66	.69	.21	.30	.50	.48
12	.48	.57	.38	.50	.62	.49
13	.74	.42	.46	.60	.60	.57
14	.58	.55	.44	.46	.32	.45
15	.53	.28	.39	.36	.52	.47
16	.50	.56	.47	.53	.49	.46
17	.48	.50	.56	.50	.72	.35
18	.48	.50	.09	.58	.66	.46
19	.53	.55	.41	.39	.46	.56
20	.59	.54	.36	.53	.48	.16
21	.57	.55	.59	.64	.65	.33
22	.50	.30	.43	.59	.63	.43
23	.40	.35	.41	.45	.56	.40
24	.59	.55	.33	.49	.50	.36
25	.41	.32	.50	.17	.50	.17
26	.47	.52	.18	.34	.49	.34

TABLE VI
 ITEM DISCRIMINATION INDEX, FORMS X AND Y,
 Grade Six (N=186)

Item	P X	B X	K X	P Y	B Y	K Y
1	.46	.41	.34	.48	.44	.48
2	.40	.31	.53	.56	.54	.03
3	.52	.56	.17	.43	.69	.43
4	.46	.34	.35	.15	.49	.21
5	.37	.61	.37	.53	.53	.55
6	.54	.50	.43	.40	.41	.21
7	.39	.51	.45	.49	.56	.52
8	.36	.63	.30	.63	.58	.33
9	.49	.57	.50	.49	.57	.55
10	.53	.59	.47	.58	.62	.53
11	.57	.65	.25	.40	.64	.48
12	.45	.52	.21	.65	.52	.57
13	.54	.55	.56	.63	.43	.59
14	.63	.49	.33	.43	.36	.54
15	.41	.26	.23	.53	.61	.40
16	.43	.64	.45	.52	.58	.50
17	.35	.55	.62	.61	.77	.57
18	.46	.35	.10	.60	.52	.41
19	.61	.64	.44	.56	.50	.46
20	.58	.55	.46	.47	.62	.29
21	.39	.52	.50	.56	.75	.23
22	.48	.26	.34	.72	.62	.51
23	.46	.32	.48	.39	.58	.51
24	.60	.43	.36	.34	.51	.25
25	.37	.31	.33	.15	.32	.13
26	.48	.52	.20	.20	.57	.34

TABLE VII
 ITEM DISCRIMINATION INDEX, FORMS X AND Y,
 GRADE SEVEN (N=175)

Item	P X	B X	K X	P Y	B Y	K Y
1	.47	.31	.51	.29	.22	.52
2	.42	.62	.48	.57	.52	.14
3	.57	.58	.30	.42	.65	.56
4	.64	.48	.53	.36	.61	.18
5	.44	.70	.49	.64	.70	.53
6	.76	.57	.49	.50	.43	.33
7	.68	.69	.49	.59	.58	.61
8	.53	.70	.50	.62	.72	.47
9	.76	.73	.66	.52	.71	.57
10	.66	.74	.65	.70	.79	.55
11	.66	.76	.36	.60	.62	.62
12	.52	.79	.37	.60	.72	.56
13	.75	.75	.52	.66	.61	.63
14	.78	.78	.55	.49	.47	.59
15	.72	.47	.54	.48	.70	.68
16	.64	.72	.57	.44	.54	.66
17	.46	.59	.52	.62	.65	.59
18	.66	.67	.15	.66	.74	.52
19	.75	.68	.41	.52	.47	.58
20	.76	.82	.67	.52	.67	.46
21	.60	.73	.52	.65	.84	.48
22	.57	.39	.52	.59	.74	.58
23	.53	.52	.62	.56	.73	.59
24	.73	.46	.57	.45	.52	.50
25	-.32	.19	.11	-.22	.24	.07
26	.24	.26	.42	.25	.27	.44

TABLE VIII
 ITEM DISCRIMINATION INDEX, FORMS X AND Y,
 GRADE EIGHT (N=164)

Item	P X	B X	K X	P Y	B Y	K Y
1	.28	.05	.26	.51	.25	.63
2	.50	.75	.61	.67	.81	.20
3	.72	.63	.45	.70	.80	.63
4	.62	.50	.61	.46	.70	.28
5	.60	.85	.51	.80	.86	.68
6	.73	.71	.62	.58	.53	.49
7	.67	.81	.53	.59	.61	.68
8	.63	.84	.56	.74	.88	.65
9	.74	.79	.70	.59	.79	.77
10	.72	.82	.63	.70	.85	.57
11	.73	.85	.38	.60	.77	.69
12	.49	.76	.42	.64	.83	.65
13	.65	.78	.65	.67	.68	.72
14	.79	.78	.62	.65	.62	.77
15	.69	.47	.58	.55	.74	.65
16	.66	.80	.55	.60	.78	.69
17	.43	.64	.60	.62	.77	.70
18	.68	.67	.39	.67	.95	.56
19	.76	.78	.54	.67	.71	.70
20	.79	.87	.66	.73	.79	.56
21	.61	.89	.60	.74	.88	.45
22	.66	.43	.54	.74	.78	.61
23	.56	.61	.64	.65	.64	.69
24	.76	.66	.66	.52	.67	.49
25	-.49	.03	.00	-.53	.06	-.20
26	.35	.20	.34	.27	.34	.21

Table IX reveals the results of the fakability study. It is surprising to note that the difference between the means for the knowledge section should be so great, since knowledge would not normally be faked. A possible explanation is that the pupils worked harder when they felt a grade was involved. Although the means show an increase in each case, the increase is not consistent. This is indicated by the relatively low value of the test-retest reliability coefficients, particularly of P and B.

TABLE IX
RESULTS OF THE FAKABILITY STUDY, FORM X,
GRADES FIVE THROUGH EIGHT (N=32)

Test Section	Mean		Standard Deviation		Pearson r	t-ratio	Confidence Level
	#1	#2	#1	#2			
P	18.8	22.6	4.9	2.4	.11	4.07	.01
B	21.7	23.0	4.0	4.0	.45	1.80	-
K	14.8	16.1	5.0	4.8	.71	1.94	-

For an N of thirty-two, only the Pearson r for section B was significant at the .05 level, while the r for the K section was significant at the .01 level. The second set of directions yielded higher scores on all three sections of the test, however only section P was significant. The overall results of the fakability study indicate that section P

is fakable, and its results can be influenced by the intended purpose of the administrator.

CHAPTER IV

SUMMARY

In summarizing the results of this study, the purpose of the study must be kept in mind. This purpose was to provide information considered in Standards for Educational and Psychological Tests and Manuals to be essential or desirable which was lacking in the manual published by the author of the test. In most cases, this purpose was accomplished. A sample of students was provided and an indication was made of its size, grade levels and limitations. A fakability study was conducted which revealed that at least part of the test, specifically section P, was susceptible to faking. The mean and standard deviation were provided for each sample used in both the reliability and fakability studies. The means and standard deviations of each grade level were indicated. The reliability of both forms of the test was found to be quite adequate when compared to standardized tests. The unusual scoring system suggested by Maehr was tested and compared with a system based on a simple total of correct answers and was found to have little or no empirical justification for its use.

The major area suggested in the purpose of the study which was not developed was the establishment of group norms. Although norms could have been established for Lutheran

schools in Kansas, it was felt that such norms would be of little or no use. Unless the pupil's directions in the test manual were revised to indicate more clearly the purpose of the test and the importance of truthfulness in answering the items, such norms would be misleading, since each teacher can presently slant the purpose of the test to his own choosing. The results of such slanting are clearly indicated by the fakability findings of this study which show that section P of the test is fakable and that all three sections produced higher means when the second set of directions was read.

It is recommended that as the test and test manual are presently constructed, the Knowledge section of the test be used only as an achievement test for diagnosing weaknesses of an individual's knowledge of Biblical information in relation to members of the local group. The norms given in the manual are not suitable for proper comparison because they are vague in their categorizing of students from poor to excellent, and do not give a breakdown of these categories by grade level.

It is also recommended that the Practice and Belief sections of the test not be used in the classroom unless (1) the directions in the test manual are revised to give the pupils more specific instructions regarding honesty in taking the test, (2) questions in section P are rewritten so

they are less susceptible to faking on the part of the students, and (3) sufficient norms are provided at each grade level for which the test was written from a clearly described and representative sample of students.

Since very little has been done in the area of testing religious attitudes, it is finally recommended that this test be used towards further research and development in this area.

BIBLIOGRAPHY

BIBLIOGRAPHY

- Ahmann, J. Stanley, and Marvin D. Glock. Evaluating Pupil Growth. Boston: Allyn and Bacon, Inc., 1967.
- American Psychological Association. Standards for Educational and Psychological Tests and Manuals. Washington: American Psychological Association, 1966.
- Educational Testing Service. Technical Report for Sequential Tests of Educational Progress. Princeton: Educational Testing Service, 1957.
- Maehr, Martin J. Manual to Accompany Concordia PBK Inventory. St. Louis: Concordia Publishing House, 1965.
- Tate, Merle W. Statistics in Educational Psychology. New York: The Macmillan Company, 1965.

AN EVALUATION OF THE CONCORDIA PBK INVENTORY

by

WAYNE HERMAN WENTZEL

B.S., Concordia Teachers College
Seward, Nebraska, 1965

AN ABSTRACT OF A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

College of Education

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1969

Statement of the problem. The Concordia Practice, Belief, Knowledge Inventory (PBK) is designed to assess the outcomes of religious education. It was prepared to discover relationships between Biblical information which upper elementary pupils possess on one hand and the practices and beliefs they possess on the other.

This test was administered to pupils of grades six through eight in seventeen Lutheran schools in the state of Kansas. The initial purpose of the study was to establish norms for Kansas on the basis of the results of these administrations. The test manual was compared to items recommended in Standards for Educational and Psychological Tests and Manuals published by the American Psychological Association and found to be lacking in several areas. These areas included (1) the reporting of the universe of the sample, (2) fakability studies, (3) reporting of means and standard deviations, (4) reporting the internal consistency, (5) incomplete directions, (6) a rationale for the scoring system used, and (7) reference groups or norms. The purpose of the study was changed to providing information which was felt to be lacking.

Sample and procedure. The sample for the study was selected from approximately 95 per cent of the students who took the test in the seventeen Lutheran schools. The sample consisted of 164 fifth graders, 186 sixth graders, 175 seventh

graders, and 164 eighth graders.

The results of these administrations were placed on IBM cards and the items computed were (1) the coefficient of reliability using the split-half method for the total score of both forms of the test at each grade level, (2) the parallel form reliability coefficient for each section (P, B, and K) at each grade level, (3) the correlation between the scoring method prescribed by the author of the test and a score based on a simple total, and (4) item discrimination indices from point bi-serial correlations for each grade level.

In addition to this information, a fakability study was conducted to determine the ability on the part of the student to present a favorable picture of himself when taking the test. Separate directions were written to have the pupil in one instance feel he was being graded on the test and in the other instance feel he was to be as honest as he could be in answering the questions.

Findings of the study. The split-half reliability coefficients proved to be quite good. The range for the total score of form X on the test was from .91 to .96 while the range for form Y was .92 to .96. When these coefficients of reliability were broken down by separate sections, they were somewhat lower, but still very acceptable. The range for form S, section P was .83 to .90; for form Y, section P

.78 to .84; form X, section B .85 to .95; form Y, section B .88 to .95; form X, section K .77 to .91; and form Y, section K .79 to .91. Coefficient reliabilities comparing the scoring method prescribed by the author of the test with a simple scoring method showed a very high correlation between the two methods, generally in the area of .99. This raised a question concerning the practicality of using the more complicated method suggested by the author of the test.

The item discrimination indices revealed that generally most of the items discriminated quite highly, with the exception of two questions. The fakability study proved to be the indicator of the greatest weakness of the test. This study showed that the test, particularly section P, could be faked. All of the sections showed some increase in the mean when the directions instructed the pupils to achieve towards a grade, and the difference between the sample means arrived at for P was significant at the .01 level of confidence.

Based on the findings of the study, it was recommended that the directions to the pupils in the manual be revised, that certain questions in the test be rewritten, and that sufficient norms be provided to use as a basis for comparison.