A Follow-Up Evaluation of the Frostig Visual-Perceptual Training Program

JAMES N. JACOBS*
LENORE D. WIRTHLIN
CHARLES B. MILLER

Introduction

DURING the 1965-66 school year an evaluation of the Frostig Visual-Perceptual program was conducted.[1] Subjects were in prekindergarten, kindergarten, and first grade. In general, it was found that children taking the Frostig program did better on the Frostig Visual-Perceptual Test than did controls. Largest gains were registered by first graders. After initial differences in Frostig scores were cancelled out, no significant differences in end-of-year reading readiness scores were found among kindergarten pupils in Frostig and non-Frostig classes. Finally, subtest analyses were made and revealed that practically all the differences in Frostig total test scores noted above were due to one subtest; i.e., Form Constancy.

Readers who are not acquainted with the Frostig Visual-Perceptual program or the instrument, are referred to the original study or to the Frostig manual.[2]

The purpose of this study is to replicate [3] as much as possible the former study and to follow up the children in the study to determine the predictive validity of the Frostig tests as well as to study the cumulative effects of the Frostig program upon future reading achievement.

Specifically, this study is designed to answer the following questions:

1. Do children in prekindergarten, kindergarten, and grade one who take the Frostig program achieve better than controls on the Frostig Visual-Perception Tests? Is the benefit uniform among grades tested?

2. Do kindergarten "Frostig" children achieve higher on the Metropolitan Reading Readiness Test at the end of their kindergarten year as compared to controls?

3. What is the predictive validity of the

[1] James N. Jacobs. "An Evaluation of the Frostig Visual-Perception Training Program." *Educational Leadership* 25(4): 332-40; January 1968.

[2] M. Frostig, P. Maslow, W. LeFever, and J. R. B. Whittlesey. *The Marianne Frostig Developmental Test of Visual Perception.* Palo Alto, California: Consulting Psychologists Press, 1964. (Standardization.)

[3] Errors in generalization are common to studies of this type. Before practical decisions regarding the use of the treatment are made, replication studies should be made.

* *James N. Jacobs, Director, Program Development, Education Center, Public Schools, Cincinnati, Ohio; Lenore D. Wirthlin, Administrative Supervisor for Reading, Education Center, Public Schools, Cincinnati, Ohio; and Charles B. Miller, Director, Psychological Services, Education Center, Public Schools, Cincinnati, Ohio*

Frostig test as compared to the Metropolitan Reading Readiness Test?

4. Do Frostig trained children achieve better on reading tests as compared to controls?

5. Do children exposed to two years of the Frostig program achieve better on reading tests as compared to one-year Frostig children and controls? That is, is there a cumulative effect due to Frostig involvement in reading achievement?

The first two questions were pursued in the initial study and are partially replicated in this study. The last two questions had to be delayed until Frostig program children last year advanced in grade.

Design and Methods of Analysis

The pupils used in this study were identified as experimental (Frostig) or control. They attended five different schools and were in intact classes. For most comparisons, controls were drawn from the same schools as experimental classes. Further, the pupils may be considered "disadvantaged" since they attend schools designated as "primary target" schools under the Education Act. Education Act services, which may bias results, are considered to be present in equal force with experimental and control children. Thus, no systematic bias is expected.

In answering the first two questions, the Frostig tests and the Metropolitan Reading Readiness Test were administered to experimental and control children in May only. The initial study included beginning-of-year Frostig testing as well as end-of-year testing. This feature could not be replicated this past year. The analysis consisted simply of a two-way analysis of variance of grade by treatment using total Frostig test score as the criterion. For reading readiness comparisons only kindergarten children were tested. Thus, a one-way analysis of variance was applied. All experimental children in this phase of the study took the Frostig program one year only.

The last three questions were studied by administering the Gates-MacGinitie test to first-grade children in May and the Stanford Primary I at the end of grade two. Since the Gates produces two subtest scores, i.e., vocabulary and comprehension, separate analyses on these two tests were made. The first-grade experimental children tested with the Gates were subdivided into two groups: one group having taken the Frostig program during the first grade only; and the second group who had taken the Frostig program during their kindergarten year as well as during the first grade. These two groups were in turn compared with control children who did not have the Frostig program either year.

The question of the effects of the Frostig program on reading achievement was studied in another context. Second-grade children in 1966-67 who were involved in the initial experiment as first graders in 1965-66 were identified and compared with respect to end-of-year (May) reading achievement using the Stanford Primary I battery. The pupils used in this phase had taken the Frostig program only during the first grade last year as compared to controls who did not have the Frostig program.

Results

Question 1: Do Frostig program children achieve better than controls on the Frostig Visual-Perception Test? A summary of the Frostig test results is seen in Table 1. The children in the experimental group took the Frostig program only for the given year.

Results of the analysis of variance showed significant differences between treatments and among grades, but no significant interaction between treatments and grades. Thus, we observe an average difference of 2.73 in favor of experimental pupils on the Frostig test. In the absence of interaction we conclude further that the benefits accruing from the Frostig program on the Frostig test score are equal to prekindergarten, kindergarten, and first grade pupils.

These results bear out the initial study with one exception: significant interaction was noted in the initial study, with first-grade pupils showing significantly more gain than prekindergarten or kindergarten children. This same result occurred but lacked statistical significance. Most important, however, is determination that the Frostig program does

produce higher Frostig scores in comparison to controls. The difference noted in interaction probably is the result of the fact that a covariance analysis was applied in the initial study using Frostig pre-test scores as a control measure while this could not be done in this investigation. Since there is little reason to believe that the experimental pupils began the year at a higher pre-test level than controls, we may interpret the higher post-test scores as responding to the Frostig program.

Question 2: Do kindergarten "Frostig" children achieve higher on the Metropolitan Reading Readiness Test at the end of their kindergarten years as compared to controls? Most of the children at the kindergarten level whose Frostig scores are reported in Table 1 also took the Metropolitan Reading Readiness Test at the end of the school year. The mean raw score made on the reading readiness test for experimental pupils was 37.22 and a standard deviation of 14.4. Kindergarten control pupils had a mean raw score of 29.90 on the reading readiness and a standard deviation of 15.2. The difference of 7.32 favoring the experimental pupils was found to be significant ($t=2.63$, df. 112).

Essentially the same comparison was made in the initial study. The results showed the experimental group to be higher than control but did not reach a statistically significant difference. In view of this replication, it seems safe to say that kindergarten pupils who take the Frostig program generally will show higher Metropolitan Reading Readiness scores.

Since pupils who take the Frostig program achieve higher on both the Frostig test and on the Metropolitan Reading Readiness Test in comparison to control, it seems reasonable to conclude that there is a correlation between the two tests. The correlation coefficient obtained between the Frostig test score and the Metropolitan Reading Readiness raw score was .556 among the experimental pupils and .659 among controls. These correlations are not significantly different and reflect a moderate degree of relationship between the two tests.

Question 3: What is the predictive validity of the Frostig test as compared to the Metropolitan Reading Readiness Test? The first-grade pupils used in the initial study were in the second grade in 1966-67. The Frostig scores of these pupils obtained in September of the first grade and again in May of the first grade were both correlated against Stanford Primary I reading subtests. The Metropolitan Reading Readiness Test scores, obtained by essentially the same pupils at the end of the kindergarten year, were also correlated with the Stanford reading subtests. The results of these correlations are seen in Table 2 (see page 172).

It should be borne in mind that both Frostig tests were given on dates *closer* to the administration of the criterion reading tests than was the Metropolitan Reading

Group	Experimental		Control		Diff. (X-C)	Grade Means
Prekindergarten	$\bar{X}$	39.00	$\bar{X}$	38.53	+ .47	38.77
	(N)	(36)	(N)	(34)		
Kindergarten	$\bar{X}$	41.78	$\bar{X}$	38.69	+3.09	40.30
	(N)	(67)	(N)	(67)		
First	$\bar{X}$	48.68	$\bar{X}$	44.06	+4.62	45.04
	(N)	(78)	(N)	(61)		
Unweighted Treatment Means		43.15		40.42	+2.73	

Table 1. Average Scores on the Frostig Visual-Perception Test* Administered in May for Experimental and Control Pupils

* The scores reported are the sums of five scale scores each of which is defined as perceptual age divided by chronological age, multiplied by 10. A total scale score of 50 is equivalent to a perceptual quotient of 100.

Readiness Test. It is well known that the closer the time between test administrations, the more probable is a high correlation. Thus, one would expect lower correlation of the Metropolitan Reading Readiness Test with the reading tests scores than either Frostig test since it was given four months before the Frostig pre-test and twelve months before the Frostig post-test.

The correlations shown in Table 2 reflect the degree to which reading achievement can be predicted from a knowledge of the Frostig and Metropolitan Reading Readiness scores. The general picture shows rather low correlation between both Frostig and Metropolitan with the reading tests. Of the eighteen correlations reported, only eight are significantly different from zero and then only barely significant.

The Frostig test given at the end of the first grade appears to be a slightly better predictor than the Frostig test given at the beginning of the first grade. There seems to be no pattern of one reading subtest being better predicted than any other. There seems to be a unique pattern of slightly better than average correlation of the Metropolitan Reading Readiness with reading test scores within the control group. All three of these correlations are seen to be significantly different from zero. One wonders why the latter correlations are systematically higher than the other Metropolitan correlations within the experimental group.

One could infer that when children go through the Frostig program their reading readiness pattern is changed somewhat, resulting in lower correlation between the reading readiness score and reading achievement later on. Thus, one may interpret this to mean that the Frostig program does *something* to the reading readiness of pupils but this "something" does not necessarily operate with equal intensity for children with low or high Frostig scores. If the latter were the case, the correlations would be maintained and only the means would be affected.

It is difficult to explain the relatively high correlation (.539) between the Frostig pre-test score and the Stanford vocabulary score within the experimental group. This correlation seems to stand alone and there appears to be no good explanation as to why this particular relationship should be higher than any other. It is presumed that this higher correlation is simply a matter of chance.

Considering the fact that the Metropolitan test was given earlier than either Frostig test and considering the fact that the

	Stanford Subtest (a)	Frostig Visual-Perception Test (b)	Frostig Visual-Perception Test (c)	Metropolitan Reading Readiness Test (d)
		N = 37	N = 37	N = 28
Within Experimental Group	Word Reading	.219	.361*	.273
	Paragraph Meaning	.144	.309*	.042
	Vocabulary	.539*	.154	.197
		N = 29	N = 29	N = 25
Within Control Group	Word Reading	.342*	.164	.432*
	Paragraph Meaning	.192	.084	.442*
	Vocabulary	.044	.324*	.494*

* Significantly different from 0 at 5% level.

(a) Stanford reading tests given at end of second grade, May 1967.
(b) Frostig pre-test given at beginning of first grade, September 1966.
(c) Frostig post-test given at end of first grade, May 1966.
(d) Metropolitan Reading Readiness Test given at end of kindergarten, May 1966.

Table 2. Correlations of the Frostig and Metropolitan Reading Readiness Tests with Second Grade Reading Achievement as Measured by the Stanford Primary I Battery

highest group of correlations was between the Metropolitan Reading Readiness Test and the reading test within the control group, one may conclude that the Metropolitan tests seem to be better predictors of reading achievement than the Frostig tests.

Question 4: Do Frostig trained children achieve better on reading tests than controls? This question was pursued by studying the average reading scores of the same pupils on whom the correlations in Table 2 were based. It should be borne in mind that these reading tests were given at the end of grade two while the experimental pupils were defined as pupils having one year of the Frostig program in the *first grade* as compared to controls who did not take the Frostig program in the first grade. Summary of the results is shown in Table 3.

The data shown in Table 3 bear out what was learned previously, i.e., that there is little relationship between Frostig scores and reading achievement. Thus we note that, while the Frostig scores of the experimental group exceed those of the controls by 3.55, nevertheless the reading subtest scores are consistently in favor of the control group. Two of the three subtests show significant differences favoring the control group. In all probability the reading differences favoring controls are due to initially higher levels of achievement by the control group. Thus it is apparent that high Frostig test performance is no guarantee of higher reading achievement.

Question 5: Do children exposed to two years of the Frostig program achieve better on reading tests as compared to one-year Frostig children and controls? First-grade Frostig children in 1966-67 were divided into two groups, one group who had taken the Frostig program as kindergarteners and a second group who were beginning the Frostig program for the first time in the first grade. End of first-grade Frostig scores were compared between the two-year and one-year Frostig pupils as well as against controls. In addition, this first-grade group was given the Gates-MacGinitie test at the end of their first grade, thus permitting comparison of reading achievement of two-year Frostig pupils, one-year Frostig program pupils, and controls. These results are shown in Table 4.

Inspection of Table 4 reveals immediately that the two-year Frostig group, i.e., pupils who took the Frostig program in kindergarten *and* in grade one performed higher on the Frostig test (51.57) than did pupils who took the Frostig program in grade one only (48.68). Both of the Frostig groups in turn scored higher on the Frostig test than did the control pupils who had a mean Frostig score of 44.06. Thus, it is apparent that the Frostig program does have cumulative effects on the Frostig test score—the longer a pupil is in the program, the higher is his Frostig test score.

Comparison of the Gates reading scores shows a picture favorable to the Frostig groups. In both the vocabulary and comprehension subtests both Frostig groups scored higher than did the controls. One-year and two-year Frostig program pupils tend to score similarly on both vocabulary and comprehension while both score com-

Stanford Subtest	Experimental Frostig Group N = 37	Control Group N = 29	Diff. (X-C)	t-ratio
Word Reading	2.12	2.44	− .32	2.13*
Paragraph Meaning	2.17	2.47	− .30	2.26*
Vocabulary	1.96	2.13	− .17	1.77
Average End of First Grade Frostig Score	53.14	49.59	+3.55	

* Significant beyond 5% level.

Table 3. Average Stanford Primary I Grade Scores Obtained at the End of Grade Two and Summary Statistics for Frostig Trained Pupils and Controls

paratively higher than controls. For both subtests, however, analysis of variance showed no significant difference among the three groups on either the vocabulary or the comprehension subtests. The lack of significant difference revealed by the analysis of variance was a result of large variation within each of the three groups.

In view of the comparative achievement of Frostig and control pupils on reading tests at the end of grade two (reported earlier), one is inclined to accept the analysis of variance at face value and conclude that there is no significant difference in the achievement of Frostig program pupils and controls on reading tests at the end of grade one.

Discussion

The results of this study may be summarized briefly as follows:

1. Pupils who take the Frostig program in prekindergarten, kindergarten, or first grade tend to show higher levels of visual-perceptual performance on the Frostig test in comparison to control pupils.

2. There is a moderate degree of correlation between Frostig scores and the Metropolitan Reading Readiness scores resulting in the fact that higher scoring Frostig pupils tend to be higher scoring pupils in terms of reading readiness tests.

3. The predictive validity of both the Frostig test and the Metropolitan Reading Readiness Test is quite low using standardized reading tests as criteria.

4. Pupils who take the Frostig program seem to have no particular advantage as far as future reading achievement is concerned as compared to pupils who do not take the Frostig program.

5. Pupils who take the Frostig program for two consecutive years achieve higher on the Frostig test compared to pupils who take the Frostig program for one year, who in turn achieve better than pupils who do not take the Frostig program in either grade.

One other finding, not discussed in the body of this report, relates to subtest analyses. In the original study, the only subtest which showed a significant difference favoring Frostig pupils was the Form Constancy subtest. This finding was disturbing in that we could not explain why this should happen. In the present study, however, we found that experimental pupils in general performed higher on *all* of the five subtests than did the controls. The differences favoring the experimental Frostig group on each test were as follows: Eye-Motor Coordination, +.81; Figure-Ground, +.61; Form Constancy, +.62; Position in Space, +.82; Spatial Relations, +.35. These results are much more reasonable and highlight the importance of replication in studies of this type.

One could argue that the differences in mental ability of the comparison groups could account for any observed differences in criterion measures. In this regard, it is interesting to note that the mean IQ's as measured by the Pintner Cunningham Mental Ability Tests of the groups shown in Table 4 were 92 for the two-year Frostig group, 91 for the one-year Frostig group, and 88 for the control group. Since these mental tests were given at the end of the year or at the same time that the criterion tests were administered, one might argue that the differences in IQ were a response to the Frostig program. One could also argue that these differences in mental ability, while small, may account

Test Comparison	Two Year Frostig Group N = 30	One Year Frostig Group N = 78	Controls N = 67
Gates Reading Tests			
Vocabulary (raw scores)	28.67	27.17	24.80
Comprehension (raw scores)	18.91	19.14	16.25
Frostig (total score)	51.57	48.68	44.06

Table 4. Average Frostig and Gates Reading Scores Obtained at the End of Grade One for One-Year and Two-Year Frostig Program Pupils and Controls

for the slight advantage seen for the Frostig groups over the control groups on the reading tests.

One may ask whether it is beneficial to offer the Frostig Visual-Perceptual Training Program to primary-grade children. If one's goal is simply to increase the visual-perceptual quotient of children, the answer is obviously "yes" since the program does tend to increase visual-perceptual scores. If, however, the goal is to increase reading achievement beyond the normally expected, the evidence obtained from this study would not tend to support the notion that this will occur. Whether reading criterion tests were administered at the end of the first grade or at the end of the second grade, the results were consistent, showing that experimental Frostig pupils achieved approximately the same as controls. While there is some *inclination* for Frostig pupils to show higher reading achievement at the end of grade one, this tendency is counterbalanced by the fact that control pupils achieved higher than Frostig pupils on reading tests administered at the end of grade two. □

A Call for Papers

JAMES RATHS *

THIS RESEARCH SUPPLEMENT is *not* designed for publishing reviews of research, discussions of research issues, calls for needed research analysis, or analysis of widely quoted research studies. *It has been established for the reporting of data.* The criteria for judging which articles are to be included in the limited space are listed in a hierarchical sequence as follows:

1. The manuscript must report data. Included in the article must be some evidence to support the reliability of the measures used in the study.

2. The article should concern itself with the behaviors of teachers (or their surrogates) and that of students as dependent variables. Behavior is taken to mean achievement scores, responses to questionnaires, etc.

3. The article should present a discussion of the results in such a manner that the meaning of the research is clear to readers. Some suggestions to meet this criterion include: a discussion of threats to the validity of the study's conclusion; an unambiguous definition of the independent variable; a distinction between the findings (data) of a study and the conclusion pertaining to the research hypotheses; a distinction between testing research hypotheses grounded in theoretical frameworks and answering research questions for which there exists no known theoretical base; and finally, establishing a basis for qualified conclusions.

Authors are invited to submit manuscripts to James Raths, Bureau of Educational Research and Field Services, College of Education, University of Maryland, College Park, Maryland 20742.

Manuscripts are welcome in all lengths, from 500 words to 8800 words. (Most articles in *Educational Leadership* are roughly 1500 words in length.) The authors of manuscripts are invited to use the style exemplified by the regular articles found in *Educational Leadership* of citing references at the end of every chapter in alphabetical order. All manuscripts will be submitted to panels of referees drawn from the membership of the ASCD Research Council and prompt decisions will be made regarding the publication of manuscripts. It is hoped that researchers in the field will begin to consider submitting reports of their work to the *Research Supplement* and that leaders in the field of curriculum development and supervision will soon direct their attention to the researches reported therein. □

**James Raths, Director, Bureau of Educational Research and Field Services, College of Education, University of Maryland, College Park*

Copyright © 1968 by the Association for Supervision and Curriculum Development. All rights reserved.