

## EVALUATION OF THE ACADEMIC ACHIEVEMENT OF SIXTH GRADE PUPILS IN A BILINGUAL EDUCATION PROGRAM

Beulah Cypress

*University of Miami\**

### SUMMARY

In 1963 Dade County established a bilingual curriculum in an elementary school. Among other goals was attainment of equal proficiency in reading in Spanish and English. Results of language and achievement testing show that while many of the goals were achieved the proficiency in English as a second language excelled that for Spanish as a second language.

In September 1963 the Dade County Board of Public Instruction established a bilingual curriculum in an existing elementary school. Coral Way Elementary School was selected since the middle class neighborhood in which it was located provided almost equal populations of North American and Cuban families. The goals of the program over a six year period included, among others, the objective of academic achievement in subject matter studied in a second language in proportion to the achievement in the pupil's native language; and the attainment of equal proficiency in reading in native and second languages.

From 1966 to 1968 an evaluative study was made of progress in language arts, arithmetic, and proficiency in second language. The findings, reported by Mabel W. Richardson (1968), indicated that the program was as effective in promoting achievement in the specific areas as the regular monolingual school curriculum, for both native English speaking and native Spanish speaking students. However, pupils from the two cultures were not as proficient in their second languages as in their native languages.

At the time of this study the highest grade attained by the children in the program was the fifth. In the spring of 1969 the school included, for the first time, a group of sixth grade pupils who had completed six years in the bilingual program. It was now important to know how the program affected this unique group.

\*Now at Florida State University

Did the Cuban children function as well in English as in the mother tongue? Did the English speaking children gain proficiency in Spanish? Were gains in the second language equal for North American and Cuban pupils? Was there any difference in achievement for the bilingual groups from children enrolled in a regular school program? Parents and school authorities were equally interested in the effects of the program on children who had spent all of their elementary school years in a bilingual education program. The experiment also presented implications for the future of bilingualism in American educational systems.

### Subjects

In the spring of 1969 certain sixth year bilingual pupils were identified at Coral Way Elementary School. There remained 17 North American and 19 Cuban pupils who had enrolled in the first grade at the inception of the program. These were the only known participants in the experimental program, and they were considered the samples for the experimental groups.

Subjects for control groups were selected from sixth grade North American and Cuban students attending a neighboring elementary school where the general socioeconomic level was judged to be similar to that of the experimental school. Due to the size of the sample no attempt was made to equate the groups on intelligence measures; however, selection for the control groups was made randomly. Thirteen North Americans and 19 Cubans made up the control groups.

### Curriculum

The regular Dade County curriculum was followed in both schools. The only difference was that the experimental groups studied approximately half the school day in native languages and the other half in second languages, while instruction in the control school was in English only. At the bilingual school, original concepts were presented in the native language by a native speaking teacher. Follow-up and reinforcement instruction was given in the second language. In some subject areas parallel tests in English and Spanish were used. (For a detailed description of the organization of the bilingual school, see Gaarder, 1967).

### Design

The following hypotheses were tested:

1. There is no difference in reading in the native language and in the second language for sixth grade native English speaking pupils who have been in the bilingual program six years.
2. There is no difference in reading in the native language and in the second language for sixth grade native Spanish speaking pupils who have been in the bilingual program six years.
3. There is no difference in achievement in a second language between North American and Cuban sixth grade pupils who have been in the bilingual program six years.
4. There is no difference in achievement in language arts between English speaking pupils who have been in the bilingual program six years and sixth grade English speaking pupils enrolled in a regular school program.
5. There is no difference in achievement in arithmetic between English speaking pupils who have been in the bilingual program six years and sixth grade English speaking pupils enrolled in a regular school program.
6. There is no difference in achievement in language arts between Spanish speaking pupils who have been in the bilingual program six years and sixth grade Spanish speaking pupils enrolled in a regular school program.
7. There is no difference in achievement in arithmetic between Spanish speaking pupils who have been in the bilingual program six years and sixth grade Spanish speaking pupils in a regular school program.

No hypotheses were developed regarding oral proficiency in native and second languages since valid tests, other than teacher judgment, were not available.

It was postulated that after six years of bilingual training, pupils would be able to read equally well in both languages; and that gains in the second language would be the same for North American and Cuban students. It was also hypothesized that

achievement in language arts and arithmetic would show the same progress for bilingual and monolingual students, thus indicating that bilingual education is not a barrier to normal progress.

### **Instruments**

The Cooperative Inter-American tests in English and Spanish were administered to the experimental groups. These tests (H. Manuel, University of Texas) have parallel English and Spanish editions with the same content and may be used for a comparison of performance in the two languages. The test in the second language was administered first, and the test in the native language followed on another day.

The Stanford Achievement tests in language and arithmetic were given to experimental and control groups. The subtests consisted of:

#### **Language Arts:**

- Word meaning
- Paragraph meaning
- Spelling
- Language

#### **Arithmetic:**

- Arithmetic computation
- Arithmetic concepts
- Arithmetic application

In order to obtain a meaningful interpretation of achievement in the two subject areas, rather than isolated measures of performance in the subtests, multivariate tests of significance using Wilk's Lambda criterion were applied to the raw scores yielded by the various subtests. Since there are overlapping concepts measured in the Stanford Achievement battery, it was not possible to add them together to obtain one measure.

The California Test of Mental Maturity had been routinely administered to all students in the two schools in the fifth grade. IQ scores which had been recorded in the pupils' cumulative folders were retrieved. These were considered as reliable and valid pre-measures where initial differences of intelligence were not controlled. (Ferguson, 1966)

It was not possible to test the students in other parts of the elementary curriculum since valid tests in these areas were not available.

### Statistical Tests

Scores on the Cooperative Inter-American tests were used to test the equality of native and second languages for the experimental groups. Since measures of language proficiency, both native and second, were obtained from the same individuals, a difference score was calculated and a t-test applied to test the significance of the difference. (a method of finding the significance of the difference between two means for correlated samples is described in Ferguson, 1966, and Hays, 1963.) One test was applied to the means of the native and second languages of the North American experimental group, and a second test to the difference between the means of the Cuban experimental group.

Comparison of the second language achievement of the North American and Cuban experimental groups was made by analysis of covariance. Since the groups were not randomly chosen, analysis of covariance was the most appropriate method. Evans and Anastasio (1968) indicated that the treatment should be independent of the covariate. Therefore it is appropriate to use the I.Q. measures as determined by the California Test of Mental Maturity if it is assumed that the Cubans were not verbally handicapped in taking the English version test.

As previously stated, the objective was to measure achievement in two specific areas, language arts and arithmetic. To accomplish this, the Stanford Achievement tests were administered but the results of the various subtests were submitted to multivariate analysis in order to arrive at the significant differences in the major subject areas. Dean Clyde's computer program, MANOVA, (Clyde, 1969) yields a multivariate analysis of covariance with adjusted means. The data were submitted to this program in the IBM 360 computer at the University of Miami. Criteria consisted of the seven subtests of the Stanford Achievement battery, four for language arts and three for arithmetic. IQ scores obtained from the California Test of Mental Maturity was the covariate.

### Results

Table 1 indicates the results of the t-tests comparing native and second languages for the two experimental groups. Comparison of the difference of the language means for the North American experimental group yields a t-ratio of 12.1 for 16 degrees of freedom, this is highly significant. The t-ratio of .085, with 18 degrees of freedom, reveals no significant difference for the Cuban experimental group.

**Table 1**  
*Comparison of Native and Second Languages*  
*Bilingual Students*

	North American (N=17)		Cuban (N=19)	
	NATIVE	SECOND	NATIVE	SECOND
Mean	110.4	74.1	101.5	101.3
Std. Dev.	11.98	18.44	13.54	17.00
t-ratio	12.1		.085	
d.f.	16		18	
Prob.	Less than .05		Greater than .05	

Results of the two analyses of covariance to test the significance of the differences between North American and Cuban experimental groups on native and second language achievement indicates that there is no significant difference between North American and Cuban experimental groups on *native* language achievement as measured by the native language versions of the Cooperative Inter-American tests. In other words, both groups of students at the bilingual school read equally well in their native languages, with the intelligence factor controlled by analysis of covariance. The results for the comparison of the same two groups, using the same covariate, as measured by the *second* language versions of the same tests showed that

there is a highly significant difference ( $p < .001$ ); and examination of the adjusted means reflects the superiority of the Cuban experimental group in second language achievement.

Multivariate tests of significance applied to the language arts and arithmetic raw scores of the Stanford Achievement tests attained by the experimental and control groups are described in Table 2, while the univariate F tests are listed in Table 3. Table 4 indicates the adjusted means of all subtests for both groups.

**Table 2**  
*Analysis of Covariance*

Multivariate Tests of Significance Using Wilk's Lambda Criterion  
Comparing North American and Cuban Experimental and  
Control Groups, Language Arts & Arithmetic

	NORTH AMERICAN		CUBAN	
	<i>Language Arts</i>	<i>Arithmetic</i>	<i>Language Arts</i>	<i>Arithmetic</i>
F	4.918	.645	2.772	1.642
d.f. hyp.	4	3	4	3
d.f. err.	24	25	32	33
p less than	.005	.593	.044	.198
R	.671	.268	.507	.360

The F-ratios in Table 2 reflect no significant difference between the Cuban experimental and control groups in arithmetic; however, they do differ significantly in language arts. The North American experimental and control groups also differ significantly in the language arts area, but not in arithmetic.

**Table 3**  
F-Tests Comparing North American and Cuban Experimental and Control Groups, Sub-Tests of the Stanford Achievement Battery. Language Arts and Arithmetic

	NORTH AMERICAN		CUBAN	
	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>
<b>Language Arts:</b>				
Word Meaning	1.472	.236	.436	.513
Paragraph Meaning	2.685	.113	.091	.765
Spelling	11.385	.002	.117	.735
Language	.946	.339	7.547	.009
<b>Arithmetic:</b>				
Computation	.093	.763	1.420	.241
Concepts	.022	.883	5.198	.029
Application	1.277	.268	1.511	.227

**Table 4**  
*Adjusted Means, Experimental and Control Groups*

	NORTH AMERICAN		CUBAN	
	<i>Exper.</i>	<i>Control</i>	<i>Exper.</i>	<i>Control</i>
<b>Language Arts:</b>				
Word Meaning	29.592	26.687	23.263	24.579
Paragraph Meaning	43.171	39.392	37.736	37.001
Spelling	39.580	28.780	31.509	32.439
Language	91.609	88.127	83.030	90.812
<b>Arithmetic:</b>				
Computation	21.370	20.824	21.132	18.973
Concepts	19.076	18.824	17.448	14.447
Application	24.155	21.721	20.740	18.681



When a multivariate test is significant it is customary to examine the univariate F-tests to determine the areas which have contributed to this significance. Inspection of Table 3 shows that the significant difference between the North American experimental and control groups for language arts lies in the spelling area; while, the language arts difference for the Cuban experimental and control groups may be attributed to language. The meaning and direction of these areas of significance will be included in the discussion.

### Discussion

Research is an ancient and revered science which imposes certain requirements so that the results of experimentation may be rendered in a form which is both interpretable and generalizable. The art of evaluation cannot yet approach the rigor of pure research. The process of evaluation of public school projects precludes strict adherence to these requirements due to the nature of the physical setting of the project and the non-maneuverability of the subjects. In education one cannot provide true experimental subjects and conditions, like rats in a laboratory, and apply a treatment. Therefore, research designs involving students and teachers are generally post hoc in nature.

Pursuit of educational excellence demands that innovative ideas be given trials, and grants are made for pilot projects on the bases of untried hypotheses. Evaluation of new programs has become an essential aspect of innovation, and it has become necessary to develop evaluation as a quasi-research method. The American school systems which are crying out for improvement cannot wait for pure research to affix the stamp of approval. True, generalization is a more valid assumption under pure research conditions, and evaluation cannot yet claim the same application of its results; but it does answer some important questions, such as: Did it work for some? How certain are you?

On the basis of answers, further application of the innovative methods can be attempted. The difference between the research-generalization corollary and the evaluation model is that research generates a probabilistic diffusion; whereas, evaluation provides a foundation for future growth and experimentation. Should weakness appear, continuance of the evaluated project may be abandoned or delayed contingent upon modification and rational

decision making. Too many successful and advantageous programs would never have been instituted had they been required to await pure research methods.

This is not to say, however, that caution should be abandoned. Assumptions underlying statistical tests should be met, and proper testing procedures applied. Threats to the validity and reliability of testing should be avoided.

In the case under consideration care was exercised to select a control group which was matched on cultural and socioeconomic variables. The experimental groups were truly involved in an experimental situation, although the samples were not selected randomly from a larger population. The subjects were the "remains" of this population. However, over the six year period little change occurred in the composition of the school and the makeup of the neighborhood. Culturally, racially, politically and intellectually it remained substantially the same.

To compensate for initial differences in "intelligence" and lack of randomization, analysis of covariance was used. Elashoff (1969) and Cronbach and Furby (1970) provide good background for the use of this technique.

While it was not possible to assign teachers randomly to treatments, it was noted that all teachers were qualified and certified, that native English speaking and native Spanish speaking teachers had common planning periods, and that teachers were proficient in their native languages. The basic Dade County curriculum was followed in the experimental and the control schools.

Tests were administered under the most normal and natural conditions possible, some of them part of the regular county-wide testing program.

*Hypotheses 1 and 2:* The test comparing reading in native and second languages within each of the experimental groups reveals a significant difference for the North Americans, while there is no significant difference for the Cuban experimental group. An examination of the raw scores shows a striking similarity between native and second languages for the Cuban children. Correlation coefficients for English language and Spanish language measured for the two groups are:

Native English speaking,  $r = .75$

Native Spanish speaking,  $r = .89$

*Hypotheses 3:* The F test for second language shows a significant difference between the North American and Cuban students in the experimental school, in favor of the Cubans. In view of the fact that the F test for native language between the same two groups is non-significant, it is clear that the Cuban students made greater progress in the second language than the North American students.

*Hypotheses 4 through 7:* Multivariate tests indicate no significant differences in arithmetic achievement between experimental and control groups. Adjusted means in the three areas of arithmetic are higher, but not significantly, for the experimental groups.

Multivariate tests, using Wilk's Lambda Criterion, in language arts indicate significant differences for both North American and Cuban students. As indicated in Table 3, the difference for the North American experimental and control groups may be attributed to the area of spelling. Table 4 shows that the mean of the experimental group in spelling is higher than that of the control group. This difference was large enough to create an all-over significant difference in the multivariate test. Table 3 also reveals a significant difference for the Cuban experimental and control groups in language arts. Again, inspection of the means in Table 4 reflects a highly significant difference in the language means. In this case, however, the mean of the control group is higher than that of the experimental.

### Conclusions

It appears that the Cuban children at Coral Way Elementary School have attained equal reading proficiency in Spanish and English. The previous study (Richardson, 1968) showed that gains were being made, but it was not until pupils reached the sixth grade that equal proficiency was reached in reading in both native and second languages.

The North American children, on the other hand, did not reach this level by the sixth grade. It must be pointed out, however, that the Cuban pupils were immersed in an English language environment outside of school. The correlation of .75 for the English speaking children might well predict the same outcome as the Cuban children ( $r = .89$ ) were they in a Spanish language milieu.

A general conclusion that by the sixth grade the Cuban children at the subject bilingual school were equally proficient in reading in two languages is tenable. While generalization under the described circumstances is not sound, it is highly probably that repeated replication will vindicate these conclusions.

The statistical result indicating greater gains in second language for the Cuban pupils is confounded by the variable of environmental conditioning. It is safe to assume that this condition will persist and that English speaking children in a bilingual school will make gains in second language learning, but may not reach the level of attainment of the new citizens. Additional studies should be pursued to compare the bilingual method against other second language methods in so far as American school children are concerned. It is possible that the gains reflected in this evaluation may prove superior to other methods despite the fact that the Cuban children exhibited superior facility in learning English.

The bilingual education program at the bilingual school did not appear to influence normal progress in language arts, or in arithmetic for the participants. Except for the language subtest result for the Cuban groups, significant differences favored the experimental groups.

The language subtest of the Stanford Achievement test measures capitalization, punctuation, dictionary skills, and sentence sense. These skills emphasize the mechanics of language as opposed to comprehension. The Cuban experimental group exhibited a weakness in the former area. The reasons for this weakness have not been analyzed. It is possible that differences in teacher effectiveness, teaching methods, time allotted for study, emphasis, etc. are responsible. It is also possible that the bilingual school emphasized oral and aural methods at the expense of the mechanics. Since the subjects who reflected a weakness in the mechanics, the Cuban experimental group, were also the subjects who showed greater gains in reading the second language, it is possible that the skills needed to write the language were neglected. These findings should be utilized to improve instructional methods for groups of the same type.

Tests are presently underway to compare progress in these and other subject areas at the junior high school level where the bilingual program has been extended. It is hoped that these studies plus past studies will present positive implications for bilingual education, not only where expediency demands it as in the case of Dade County where the school population includes almost 11% Cuban children, but where the advantages of bilingualism will create a permanent place for this type of instruction in the American educational systems.

### References

- Clyde, D. *Manova*. Program distributed by Clyde Computing Service, 1969. Box 166, Coconut Grove Station, Miami, Florida 3313.
- Cronbach, L. J. and Furby, L. How we should measure change - or should we? *Psychological Bulletin*, 74, 1970, 68-80.
- Elashoff, J. D. Analysis of covariance: A delicate instrument. *American Educational Research Journal*, 1969, 6, 383-402.
- Evans, S. H. and Anastasio, E. J. Misuse of analysis of covariance when treatment effects and covariate are confounded. *Psychological Bulletin*, 1968, 69, 225-234.
- Ferguson, F. A. *Statistical analysis in psychology and education*. New York: McGraw-Hill Book Company, 1966.
- Gaarder, A. B. Organization of the bilingual school. *Journal of Social Issues*, 1967, 23, 110-120.
- Hays, W. L. *Statistics*. New York: Holt, Rinehart and Winston, 1963.
- Richardson, M. W. An evaluation of certain aspects of the academic achievement of elementary pupils in a bilingual program. Unpublished Ed.D. Dissertation, University of Miami, 1968.