DOCUMENT RESUME

ED 086 423 56 RC 007 574

AUTHOR Orvik, James M.
TITLE Final Evaluation Report: Primary Eskimo Project of the U.S. Bureau of Indian Affairs, Bethel Agency, Bethel Alaska, July 1973.
INSTITUTION Alaska Univ., Fairbanks. Center for Northern Educational Research.; Bureau of Indian Affairs (Dept. of Interior), Bethel, Alaska. Bethel Agency Office.
SPONS AGENCY Bureau of Elementary and Secondary Education (DHEW/OE), Washington, D.C. Div. of Bilingual Education.
REPORT NO BIA-EOOC14200759
PUB DATE Jul 73
NOTE 40p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS American Indian Languages; Basic Skills; *Bilingual Education; *Early Childhood Education; English (Second Language); *Eskimos; Federal Programs; *Language Ability; *Program Evaluation; Skill Development; Statistical Analysis; Tables (Data)
IDENTIFIERS Alaska; BIA; Bureau of Indian Affairs; Yupik Language

ABSTRACT

The Bethel Agency, Bureau of Indian Affairs, has now completed its 3rd year of experimental bilingual education. Beginning in 1970 with 3 pilot schools and adding 4 schools the following year, 3 grade levels in 17 Kuskokwim River village classrooms have been reached by the Primary Eskimo Program. To determine the validity of this approach, an evaluation design was conducted over a 3 year implementation period. The 2 categories of instruments used for the present evaluation were: (1) academic--Yupik literacy and numerical skills; (2) linguistic--acquisition of grammar and meaning in Yupik and English. The analysis procedure used was the method of t-test for differences between independent group means. The results of the statistical analysis are presented in 3 main sections: (1) literacy skills, (2) numerical skills, and (3) linguistic skills. In each section, the results are generally described, followed by a brief discussion of the overall patterns as a whole. Program recommendations in the final section of the report cover: (1) instructional objectives, (2) math, (3) alphabet, (4) creative writing, and (5) English as a second language. (FF)

Final Evaluation Report: Primary Eskimo Project of the U. S. Bureau of Indian Affairs, Bethel Agency, Bethel Alaska

July 1973

by

James M. Orvik, PhD.
Project Evaluator
Center for Northern Educational Research
University of Alaska
Fairbanks, Alaska 99701

This report is funded under the U.S. Bureau of Indian Affairs contract EOOC14200759. The opinions expressed are those of the author and do not necessarily reflect the views of the U.S. Bureau of Indian Affairs or the U.S. Office of Education.

INTRODUCTION

The Bethel Agency of the U.S. Bureau of Indian Affairs has now completed its third year of experimental bilingual education. Beginning in 1970 with three pilot schools[1] and adding four schools the following year, children in seventeen Kuskokwim River village classrooms, spanning three grade levels, have been reached by the Primary Eskimo Program (PEP).

In an area where Yupik Eskimo is often the first and only language of its native citizens, a decision was made to experiment with teaching the subject matter of the early primary grades in the Native language, introducing English as a second language in small, manageable portions. This approach contrasts with traditional education in the remainder of the Bethel agency schools in which virtually the entire school day is taught in English.

In order to determine the validity of such an approach an evaluation design was established to be conducted over a three year implementation period. The goal of this evaluation was to assess by objective means the overall impact of the program in selected performance domains. Evaluation in the first two program years focused on determining the extent to which general language changes were taking place in the bilingually taught children. Since vocabulary growth is an important index of such changes tests were developed to measure Yupik and English vocabularies of children in the newly implemented programs. These scores, obtained at the beginning and end of the school year, were compared with those obtained in comparable village schools

---

[1] The 1970 program was begun in first grades only, adding a new level with each succeeding year.

nearby in which traditional programs were being carried out.

It was found by the end of the school year that children in the bilingual schools had gained not only more Yupik vocabulary as was expected, but had also gained significantly more English vocabulary, even though English instruction comprised only a fraction of the school day.

In the second year of the program, the number of schools participating at the first grade level was expanded to seven, and the original three schools added second grade bilingual classrooms. Again, relative vocabulary gains comprised the main evaluation focus but this time with an outcome somewhat disconcerting to program advocates. While the first graders showed the same significant gains shown by first graders the previous year, the second grade students did not perform as well as their counterparts in comparison schools. This reversal led to a good deal of speculation regarding the general efficacy of the program, the adequacy of the testing program and other aspects of program implementation. For a number of reasons, the second year reversal was concluded to be the result of several factors, chief among which were (a) the wearing off of the initial novelty of the approach, and (b) the problems of preparing for and adding a second program level. The evaluation design for the third year has data relevant to these questions, and among other things, examines whether the second level problem is a result of its having been a new addition or if there are more deep-seated questions raised about the validity of the bilingual approach in general. If the second grade in this year's program shows relatively poor performance, real cause for concern might be warranted. On the other hand, if there is relative improvement in the second level followed by poor performance at the third level, other theories of explanation are in order, one of which may be that the addition of a new level will always presage a transitional setback pending the passage of time and gains in experience.

Needless to say, the first two years of the bilingual experiment demonstrated a mixed performance picture from which only incomplete conclusions could be drawn. Whether the momentum in language gains consistently shown by the first graders holds for the upper primary grades is the focus of a broadened third year evaluation design reported presently.

Past Evaluation Designs

The reader should bear in mind that the main features of the evaluation activities have been shared by both of Alaska's education agencies conducting bilingual programs in Southwestern Alaska; the Bureau of Indian Affairs (BIA) and the Alasl. State Operated School System (ASOSS). This arrangement has allowed a broader based opportunity to establish comparison groups from which to obtain comparative data. The procedures (though not the results) described here represent cooperative effort between the two agencies which has evolved over the past three years of program operation.

As noted above, the first and second year evaluations included a focus on general language changes in the bilingually taught children. In addition, measures of non-verbal intelligence were used to asses if other domains of performance could mediate changes in the behavior of the children. It was soon discovered that additional testing time involved with the latter measures did not warrant their continued inclusion in the evaluation design because their adaptation to a cross-cultural setting left them open to ambiguous interpretations.

The first year testing was conducted by a trained testing team who traveled to each bilingual and comparison village. In the second year, however, testers were trained in each of the program villages, eliminating the travel expenses incurred by a testing team. The comparison data for the second year consisted of data obtained in each program village from students in the

upper primary grades, not yet included in a bilingual program.[1] In most ways, this procedure came to be considered unsatisfactory from the standpoints that (a) objectivity was somewhat compromised by having teachers test their own students, and (b) diffusion of bilingual program effects from the lower to upper grades could not reasonably be controlled for statistically.

The evaluation design for the present year has thus evolved from substantive as well as logistical findings of the first two years and has, therefore, changed in major ways. First, the number of performance skills measured has been expanded to reflect the need for specific information sought by program officials. Second, sources of objectivity compromised in the second year evaluations have been restored by returning to the testing team concept. Third, rather than attempt to test all of the children in seventeen target villages (ASOSS, BIA and comparison) a stratified random sample has been selected to minimize the loss of classroom instruction time for testers as well as students. And fourth, the testing has been limited to a single post test period, since the degree of initial comparability of comparison and bilingual schools had been satisfactorily established in the prior evaluation years.

THIRD YEAR EVALUATION DESIGN

Instruments

The instruments used for the present evaluation fall into two main categories, (1) academic -- Yupik literacy skills, and numerical skills; (2) linguistic -- acquisition of grammar in Yupik and English and acquisition of meaning in Yupik and English.

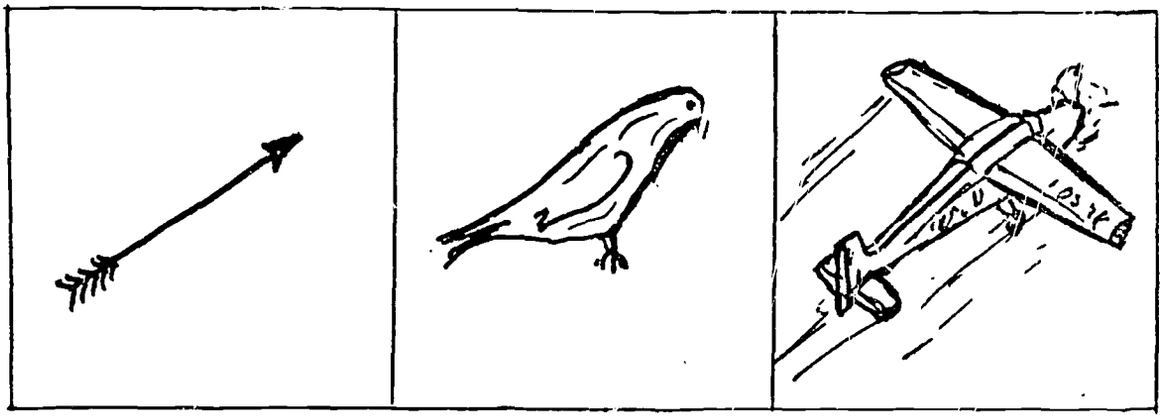
---

[1] Comparison data for the first grade children consisted of data gathered in the comparison schools the previous year.

Academic Instruments--Yupik Literacy Skills

The measurement of Yupik literacy skills was divided into three main catagories: (1) prereading, (2) decoding, and (3) encoding skills. The measurement of prereading skills consisted of the (a) recognition of initial letter sounds using stimuli such as:

P



and directions (in Yupik):

Examiner points to initial letter "p" and says:
"This makes the sound "p".

The examiner then points to each of the three pictures and says:
"One of these things begins with the sound "p"; which one is it?"

If the child points to the correct one (no. 1) say:
"Very good! "pitegcaun" begins with the sound "p".

If he doesn't point to the correct one, point to the one he chose and say:
"What do you call that?"

If what he calls it begins with "p", say:
"Very good! And this one (pointing to pitegcaun) also begins with "p", doesn't it?"

If it doesn't, go back to the beginning, and this time name the pictures for him. Do this as often as needed for the child to get the idea.

Scoring:

The number of pictures the child selects correctly;

followed by similar stimuli and appropriate directions, (b) the visual discrimination of symbols using the stimulus:

Demonstration:

at atam

and items:

pi	tupik	tupa	qupellrua
ta	tema	tuma	atata
mu	maraq	murak	nutek

with directions:

Line up the words to the child's point of view.

Point to demonstration item "AT", and say:
"These letters can be found over here. Point to the place over here where you can see these letters together."

If the child points appropriately, say:
"Very good! You found them."

then hand the child a pencil and say:
"Now make a circle around the letters."

If he doesn't point appropriately, you point them out for him, give him a pencil and say:
"Now make a circle around the letters we found."

Go on to the next one; "PI, "TA", and "MU".

and (c) reading phonemes using the stimuli:

c	y	r	n
v	ng	k	q
i	u	e	

and directions:

Point to each phoneme and say:
"What sound does this make?"

Scoring:

Write on the answer sheet a plus if the child says it correctly, or the sound the child makes is incorrect.

Stop after three failures in a row.[1]

The measurement of Yupik decoding skills consisted of (a) reading sight words with the stimuli:

tai-tai	una	neri	ayii
waniwa	atraa	qimugta	
kuimaa	mayua	ermia	

and directions:

Point to TAI-TAI in the book and say:
"Here are some words from your reading books. Let's see how many you know. Start with this one."

Scoring:

Write down the word the child says or a plus if he gets it right. Stop after three failures in a row.

(b) Decoding new words, using the stimuli:

patu	amci	qilak
qalu	qamiquq	uqilauq

and directions:

Point to "patu" in the booklet and say:
"Here are some new words. How many can you figure out? Start with this one."

Point to and sound out "patu."

Go on to the rest of the words and give no further help.

Scoring:

Write down the child's decoding attempts or a plus if it is correct. Stop after two failures in a row.

(c) matching words with pictures, (see Appendix for stimulus material) with the following directions:

---

[1] In the latter case, children in comparison schools were given full credit even if they gave an appropriate English response to a particular phoneme. In general, instructions were appropriately modified to adapt to the comparison school environment.

Point to "ENA" on the answer sheet and say:
"This word says ENA and here is a picture of ENA. See how I draw a line to connect them. These other words have pictures too. Draw a line to connect the words to each picture."

and (d) reading and following simple directions, with the stimuli:

Demonstration: quuyurni

aqumi natermun	qengan enirru
amiik ikiresgu	put'en
qeckaa	unaten elliu qamiquvnun

and the directions:

Point to "quuyurni" in the booklet and say:
"This says to do something. It says "quuyurni." (demonstrate for him).

Point to the rest of the directions and say:
"Each of these says to do a different thing. Read each one out loud and do what it says to do. Start with this one." (Point to the next direction).

Scoring:

Score 0 if he doesn't read the direction.
Score 1 if he reads it but doesn't do what is asked.
Score 2 if he does what is asked.

Stop after he misses two in a row.

Thus, the measurement of decoding skills is designed to assess not only simple decoding performance as in (a) and (b), but also his ability to attach recognition meaning to the symbols he decodes, as in (c) and behavioral meaning as in (d).

Finally, the measurement of encoding skills assesses three levels of written performance: (d) ability to write the alphabet (appropriate to Yupik or English), (2) ability to encode Yupik sounds and words, and (3) performance in which the pupil writes about himself.

In writing the alphabet the following directions were used:

Point to the place on the answer sheet where the child is to write the Yupik alphabet and say:

"Here is some paper to write the letters of the Yupik alphabet (or English if control school). You know A, C, E. (or A, B, C, for English). Start here and write them all."

Scoring:

1 point for each correct letter (appropriate to the alphabet the child attempts; Yupik or English).

0 points for each repetition.

0 points for reversed or poorly aligned letters.

Divide the total points by 15 (Yupik) or 26 (English) and multiply by 100 to get the per cent correct.

The next level of measured encoding skills consisted of the child's ability to encode correctly Yupik sounds (phonemes) and words given the following stimuli:

1. pa	8. nu	15. nuna
2. pu	9. una	16. patu
3. tu	10. ani	17. unan
4. ta	11. ata	18. tauna
5. ci	12. ayii	19. nutek
6. ca	13. qaku	20. camek
7. ni	14. tuma	21. panik

and directions:[1]

Point to the place on the answer sheet for writing sounds and words and say:

"I want you to write down (point) here the sounds that I say. Let's start with "pa", etc."

Scoring:

1 point for each corrected spelling divided by 21 and multiplied by 100 to give the percentages.

Stop after four in a row are incorrect.

Finally, the highest levels of incoding performance measured in the present evaluation consisted of eliciting a written essay from the child in the

---

[1]Words properly encoded in English were counted correct in comparison schools.

following manner:

> Point to the appropriate space on the answer sheet and say:
> "Now I want you to write something about yourself. Write anything you want and write as much as you want. You could write about where you live or your family or about things you like to do."
>
> Allow 10 to 15 minutes for the child to finish, alone if possible.
>
> Scoring:
>
> > Score 1 point for each word written.
> > Score 1 point for each correctly spelled word.
> > Then divide the total number of correctly spelled words by the total number of words.

The child was allowed to respond in which ever language he was most comfortable.

As can be imagined, the initial scoring of the essays posed some problems, chiefly in making cross-language comparisons in an objective manner. The final scoring was accomplished by the following steps.

> Step One: All responses were translated into a common language--English--by experienced Yupik teachers. In all cases errors of grammar and syntax were translated as faithfully as possible. Nonsensical responses were noted as such.
>
> Step Two: All responses were then transcribed to 3x5 cards by a third party with the child's name, school, and grade level on the reverse side.
>
> Step Three: Three experienced rural Alaskan primary teachers independently rated the responses on a seven point rating scale, basing their judgements on the overall quality of the written content, disregarding spelling errors.
>
> Step Four: The independent ratings were combined to

give each child a total essay score.[1]

Academic Instruments--Numerical Skills

Measurement of numerical skills consisted of two main components. The first focused on the ability of the children to use names of numbers for counting and the second focused on the ability of the children to perform a variety of arithmetic calculations.

Counting skills were assessed in three parts: (1) oral counting, using the directions:

> Counting from 1 to 30 in Yupik (or English in comparison schools).
> Say to the child, "Now I would like you to count from 1 to 30 in Yupik (or English in comparison schools).
>
> Scoring:
>
> > Score 1 point for each correct number.
> > The score was the highest number up to 30 the child could count without error.

(2) recognition of numbers, with the stimuli:

> 8 0 2 3 7 1 5 2 9 6 4
> 56 75 42 64 57 13 35
> 178 124 281 339

and directions:

> Point to the number and say:
> "What number is this?"
>
> Same directions for the second row of numbers except discontinue after three failures in a row.
>
> And the same directions for the third row except discontinue after two failures in a row.
>
> Scoring:
>
> > Score 1 point for each correct number. (Note that two and three diget numbers are only single numbers, for example 56 is not

[1] As an estimate of the reliability of the three judges ratings, the patterns of these agreements over the 189 responses were as follows: All three agreed: 126 (66.7%); two agreed 62 (32.8%); all three disagreed 1 (less than 1%). In only two cases did disagreements span more than two scale points. For all possible judgements there was 89% agreement.

5--6, (five--six) it is fifty-six.)

and,

(3) counting objects, with the stimuli shown in Appendix A,

and directions:

> Say to the child, "Now I would like you to count these things and tell me how many of each there are. Start with these. (point to the triangles).

Scoring:

> Score 1 point for each correct answer.

The second component of numerical skills measurement consisted of basic arithmetic as follows:

> The tester writes out each problem on the child's tablet (one at a time) and says: "You take the pencil and do this one." After he does the first one, write out the second one and so on.

Scoring:

> Score 1 point for each correct answer.

Addition:

3	1	6	6	4	11	36	104
+2	+3	+3	+5	5	+ 5	+15	+ 9
				+1			

Discontinue after three failures in a row and go on to subtraction.

Subtraction:

4	5	8	8	56	41	232	332
-1	-3	-2	-4	- 5	- 7	- 21	- 25

Discontinue after three failures in a row and go on to multiplication.

Multiplication:

1 x 2 = _____

4 x 3 = _____

7 x 3 = _____

Linguistic Instruments

For the broad purpose of assessing comparative language changes in program and non-program children, two subtests of the Illinois Test Psycholinguistic Abilities (ITPA) were adapted, --The Grammatic Closure, and the Auditory Association subtests. According to Kirk et al, (1968)[1] the Grammatic Closure subtest:

> ...assesses the child's ability to make use of the redundancies of oral language in acquiring automatic habits for handling syntax and grammatic inflections. In this test the conceptual difficulty is low, but the task elicits the child's ability to respond automatically to often repeated verbal expressions of standard American speech. The child comes to expect or predict the grammatic form so that when part of an expression is presented he closes the gap by supplying the missing part. The test measures the form rather than the content of the missing word, since the content is provided by the examiner. (p. 11)

The Auditory Association subtest:

> ...taps the child's ability to relate concepts presented orally. In this test the requirement of the auditory receptive process and vocal expressive process are minimal, while the organizing process of manipulating linguistic symbols in a meaningful way is tested by verbal analogies of increasing difficulty. A sentence completion technique is used, presenting one statement followed by an incomplete analogous statement, and allowing the child to complete the second statement appropriately. (p. 10)

These two tests, as published, are designed to deal with verbal output at two different levels of language organization. By adapting the stimulus material to the familiar locale of the Eskimo child, and adapting by translation the verbal content of each test item it was hoped that similar processes would be measured in children affected by the bilingual education program. To be sure, the difficulty of achieving a perfect adaptation of

---

[1] Kirk, S.A. Mc Carthy, J.J. and Kirk, W.D., Examiner's Manual: Illinois Test of Psycholinguistic Abilities, University of Illinois Press, Revised Edition, 1968.

both the visual stimulus material and the verbal item content is great and as many steps as possible have been taken to assure appropriatences within the given situation.

The test adaptations were made in conjunction with personnel of the Eskimo Language Workshop, whose task it was to (a) modify test pictures to the local environment, (b) translate item content into meaningful tests of grammatical structures, and (c) provide back-translations for use in corresponding English language items. In most cases the English version is not a direct literal translation of the Yupik, but has been readapted to make the syntax meaningful as a test item.

The linguistic tests were administered as follows:

1. Grammatical closure--each item has a stimulus picture and oral verbal statement, beginning with a demonstration:

"Waniwa-llu malruk _____." (pointing to the picture of two beds, and gesturing for the child to complete the sentence). If the child doesn't respond he is given the answer and the item is repeated until the child gets the idea before proceeding to the first item in the 32 item set. The same procedure is used in the English version of the test.

2. Auditory Association--each item represents a verbal analogy which the child must solve by giving orally an appropriate completion to the sentence given by the administrator. The test is begun with a demonstration item: "Aataq ang'uq, piipiq _____." (father is big, baby is _____.) followed by the reverse: "Piipiq mik'uq, aataq _____." When the child shows he has the idea, the tester proceeds through the remaining 38 items.

For both the Grammatic Closure and Auditory Association tests, the Yupik and English versions are administered seperately.

Table 1 summarizes the preceding description for quick reference by the reader. Included in Table 1 are the combinations of subtest components

used in the final statistical analysis. For example, a total pre-reading score was obtained by combining its three subtest components; initial letter sounds, visual discrimination of symbols, and reading phonemes. In most cases, subtests were combined where it would ease the burden of statistical calculations, provided there was reasonable homogeniety of content. In cases where subtests are not combined it was felt the subtests either were measuring divergent skills (e.g., numerical skills) or used measurement scales too varied to permit combination without undergoing time consuming statistical scale transformations (e.g., encoding). The resulting combinations of tests and subtests summarized in Table 1 provide a total of eleven units for statistical analysis.

Testing Procedures

All tests were administered by experienced Yupik bilingual teachers recruited from the ASOSS and BIA bilingual programs. Testers were selected according to four major criteria: (1) personal interest in the testing program, (2) recommendation by principal teachers involved in the program,(3) availability for travel to a training workshop, and (4) assent by the majority of bilingual aides. Of the eight selected, three had prior experience as testers in earlier evaluation activities.

The testers received the main portion of their training at a three-day workshop held in early March at the Bureau of Indian Affairs site in Bethel. During the three days, general testing concepts such as measurement and random sampling were assimilated as well as specific administration procedures. In addition, the testers gave substantial input into the final structure of the tests, developed scoring criteria, and laid the groundwork for the math test to be used in the program.

Following the workshop, final production of the tests was completed, a final testing schedule was developed, and a random sampling plan was finalized

TABLE 1

SUMMARY OF EVALUATION TESTS BY GRADE LEVEL

Ability Tested	Instruments	Number of Subjects by Grade Level					
		Level One		Level Two		Level Three	
		Bil.	Comp.	Bil.	Comp.	Bil.	Comp.
Literacy Skills:	prereading						
	initial letter sounds						
	+visual discrimination of symbols						
	+reading phonemes	31	19	31	19	15	15
Decoding:	reading sight words						
	+decoding new words						
	+matching words with pictures						
	+reading and following directions	31	19	31	19	15	15
Encoding:	alphabet	31	19	31	19	15	15
	sounds and words	31	19	31	19	15	15
	free essay	31	19	31	19	15	15
Number Skills:							
Counting:	oral counting						
	+naming numbers						
	+counting objects	31	19	31	19	15	15
Arithmetic:	arithmetic	31	19	31	19	15	15
Linguistic Skills:							
Grammar Closure:	Yupik	31	19	31	19	15	15
	English	18[a]	19	18	19	15	15
Auditory Association:	Yupik	31	19	31	19	15	15
	English	18	19	18	19	15	15

a. For some unaccountable reason, one tester did not administer the English versions of the Grammatic Closure and Auditory Association Tests, reducing the number of subjects from 31 to 18 both in grades one and two

including the village assignments for each tester. Travel arrangements were coordinated locally by BIA and ASOSS area administrators. Appendix B contains the finalized administration instructions, testing assignments, and sampling procedures for all posttesting activities. With few exceptions the testing program was carried out satisfactorily. The few exceptions were the result of unforseable local conditions requiring on the spot decisions by the particular tester. In only one case (see Table 1, footnote) was there a significant loss of data, but even then the ability to draw data-based conclusions was not seriously jeapordized.

RESULTS AND DISCUSSION

The analysis procedure used throughout was the method of t-test for differences between independent group means.[1] To explain further, the t-test provides an estimate of the probability that two group means could differ an observed amount simply by chance. A decision can thus be made whether or not to place confidence in the effectiveness of a program. This is done by rejecting or not rejecting the idea that a particular test result, comparing a group of program with a group of non-program children, could have happened as a result of chance or luck. For example, if a t-test shows

---

[1] The t-test for independent means proceeds as follows:

$$t = \frac{\bar{X}_b - \bar{X}_c}{S_{\bar{x}_b - \bar{x}_c}}$$

Where: $\bar{X}_B$ is a mean score for a bilingual group
$\bar{X}_c$ is a mean score for a comparison group, and
$S_{\bar{x}_b - \bar{x}_c}$ is the standard error of the mean difference, estimated from the two sample variances. The resulting t value is compared with tabled t values for various sample sizes. For a more technical treatment of the t statistic the reader is referred to Edwards, A.L., Experimental Design in Psychological Research., Holt, Rinehart and Winston, New York, 1963, Chapters 7 and 8.

that a difference between two group means could be expected to happen by chance not less than twenty times in a hundred (i.e., with a probability (p) greater than .20) we would fail to reject the notion that the difference was due to chance and thus have little confidence in the idea that the program was effective. If, however, a t-test shows that the means could differ by chance fewer than five times in a hundred ($p < .05$), we will have reached the commonly accepted scientific standard for rejecting the idea of chance differences and, therefore, be able to have confidence that the program was indeed effective. Of course, the same decision rules hold for cases in which the comparison group does better then the program group. Such cases are shown in the results as negative t-test scores.

The results of the statistical analysis are presented in three main sections: (1) literacy skills, (2) numerical skills, and (3) linguistic skills. In each section the results are described generally, followed by a brief discussion of the overall patterns in the results taken as a whole. Program recommendations are given in the final section of the report.

Literacy Skills

Table 2 shows the results of the statistical analysis of performance in literacy skills for each grade level. The reader is reminded that the values for t are the best index of comparative performance since they indicate whether a particular mean difference between a bilingual and comparison group should be taken seriously, i.e., represents a significant program difference. Negative t values indicate a higher comparison group mean.

Beginning with prereading skills the performance of the bilingual program children was substantially superior. This superiority was most marked in grades one and two. By level three, both the bilingual and comparison groups were about equal but this may have been largely due to the test

TABLE 2

LITERACY SKILLS IN YUPIK

	Prereading			Decoding			Encoding								
							Alphabet			Encoding Sounds			Open Essay		
	Bil.	Comp.	t	Bil.	Comp.	t	Bil.	Comp.	t	Bil.	Comp.	t	Bil.	Comp.	t
Level One	15.20	9.70	3.20**	13.94	1.16	6.09**	65.39	70.94	-.57	31.65	1.26	4.44**	5.61	2.42	2.59*
Level Two	14.97	11.95	2.32*	17.45	1.53	6.03**	82.22	83.10	-.12	48.06	.26	5.33**	7.97	7.58	.28
Level Three	15.87	14.83	.99	18.73	4.07	4.25**	88.87	98.67	-3.45**	57.13	3.47	4.69**	8.00	14.10	-3.11**

* $p < .05$
** $p < .01$

leaving no more room for improvement. The important result is that first grade performance is high, giving evidence of a good beginning in Yupik literacy, comparable to what might be expected by the third year in the traditional program where the children must rely on their ability to generalize from what they have learned in English literacy training. In Yupik decoding skills, the bilingual program children show clear superiority at every grade level.

In encoding, a rather mixed picture of performance has taken shape. While the ability of the bilingual program to establish the concept of the written alphabet is weak, (especially at level three) the ability of the children to encode Yupik sounds and words successfully is quite strong at every level. Relative skill at free written expression begins strong in bilingual at level one and then fall clearly behind by level three. Infact, many of the level three children made no attempt to write anything at all.

The reader should bear in mind that the purpose of evaluation in this section was to assess the degree to which the bilingual classroom is able to prepare Yupik speaking children to be literate in their first language. Using the traditional classroom as an estimate of what might have happened otherwise, makes relatively clear the general success in meeting this goal. The only exceptions are in areas in which children in comparison schools were not restricted by the tests from relying on English as a mode of written expression. In all other cases, virtually no generalization from English to Yupik was in evidence by children in the traditional program. It would, of course, be unfair to say that no literacy skills in English are being developed in the traditional schools since the evidence certainly does not support such a conclusion.

It does seem certain that a concept of the alphabet is not necessary for other basic encoding operations particularly in the accurate formation

of sounds and words received aurally. The bilingual children do fine without it and the traditionally taught children are at no apparent advantages possessing it (recall that spelling was not considered in judging the free essay). Perhaps teaching an alphabet is more for the reinforcement of the teacher than of the pupil and hence constitutes an unnecessary part of the curriculum. In fact, trying to establish an alphabet concept early may only lock the child into an ungeneralizeable system which later the child is required to repudiate upon second language literacy training.

Generally it appears that a sound basis in literacy skills is being developed in the bilingual program but careful note should be taken on the tendency toward questionable performance in free written expression.

Numerical Skills

Table 3 shows the test results for assessing comparative numerical skills. Two components were tested; a component comprising counting and number identification (naming) skills, and a component comprising common arithmetic calculations. In the former, counting, the comparison groups performed significantly better than their bilingually taught counterparts at each of the three grade levels. However, in arithmetic the bilingual program children performed as well or better than the comparison school children.

The problem of establishing a Yupik math curriculum has been present since the program's beginning. First of all, there exists no standard treatment of math throughout the bilingual program schools. For example, program schools vary in the time at which English names for numbers are introduced. Second, most Yupik counting systems are developed on a metric using the base twenty, necessitating highly complex transformations into the English system of base ten. For numbers below 20 or 30, there is

TABLE 3

COMPARATIVE NUMBER SKILLS

	Math Component Scores					
	Counting			Arithmetic		
	Bil.	Comp.	t	Bil.	Comp.	t
Level One	31.00	40.08	-3.16**	2.50	2.10	.52
Level Two	37.10	50.50	-5.40**	7.40	5.10	2.09*
Level Three	34.50	56.30	-6.49**	10.70	10.90	-.10

* $p < .05$
** $p < .01$

generally no difficulty but numbers greater than 30 begin to possess long and cumbersome names, which are linguistically, as well as mathematically different from their English equivalent.

Why then should arithmetic pose no apparent problem given the counting difficulty? One possible reason is that the arithmetic problems used in the present evaluation are, like most arithmetic operations, approachable by reduction to single integers. In fact, even into secondary school most math calculations are performed by reduction into single digit operations. This may be why the children in the bilingual program can handle calculations reasonably well without the apparent facility with large number concepts. However, it follows that when such concepts become necessary at some later time the children in the bilingual program may well have problems developing the necessary abstractions to go forward. This evidence calls for serious consideration to redesigning the entire Yupik math curriculum; from materials development to teacher preparation.

Linguistic Skills

Linguistic skills in each language were measured in two ways, one stressing the acquisition of grammar and syntax and the other stressing the development of meaning.

Referring to Table 4, in Yupik, the quality of performance in grammatical use is clearly greater at grades one and two for the bilingual program students, with the trend carried, though weaker, into grade level three. In their ability to deal with meaning in Yupik the bilingual program children show significantly better performance at all three grade levels.

In English grammatic development there is a strong beginning performance by bilingual program children which tapers off in the later primary grades so by level three the bilingual students are still holding their own but

TABLE 4

COMPARATIVE LINGUISTIC SCORES

	Yupik						English					
	Grammatic Closure			Auditory Association			Grammatic Closure			Auditory Association		
	Bil.	Comp.	t	Bil.	Comp.	t	Bil.	Comp.	t	Bil.	Comp.	t
Level One	18.65	11.79	4.23**	15.32	8.32	4.05**	4.72	2.68	2.65**	5.22	5.11	.08
Level Two	19.26	14.21	3.63**	18.19	13.16	2.89**	6.11	4.53	1.34	9.06	8.74	.20
Level Three	20.73	18.60	1.06	21.60	18.20	1.95*	8.47	8.53	-.04	12.27	16.60	-2.35**

* $p < .05$
** $p < .01$

not showing the distinct advantage they began with. The development of facility with meaning in the English language is essentially equal for both groups until level three where the comparison students show clear superiority.

It should be noted that of the two English tests, Grammatic Closure measures the abilities most stressed in the English as a second language (ESL) portions of the bilingual program curriculum, and this is where the stronger performance was noted, especially in early grades. This finding may indicate the the ESL component of the program is successful but only within a limited range of intended outcomes relative to what happens within the traditional program where the weight of exposure to English apparently gains momentum by the third year or so. By this logic it is reasonable to predict that the early basis in English grammar and syntax and the gains in the language development in Yupik will show a multiple effect when increased exposure to English takes place in the post primary years.

General Discussion

The most striking pattern to emerge from the data taken in aggregate is the marked tendency for the level three performance to show a sizeable drop. The reader will recall from the introduction to the present report that the same tendency was shown last year when the present level three students were at level two. A theory was advanced earlier that the phenomenon may have been due to a lag in program development when a new level is added. To that theory may be added several others, given the evidence.

First, the earlier theory may still hold. In fact, the greater strength of performance of level two children this year lends support to this line of speculation since it now appears as though the level two curriculum has taken

shape. At the same time the newly added level, level three, shows the same evidence of tapering off that level two did when it was first added to the bilingual program design.

But a second theory also warrants consideration. The performance drop could also be due to a relatively low ability level of the particular children comprising this year's level three, and they just happen to be the vanguard of the new movement so when a new grade level is added to the program, they are "it." This always a painful area of speculation, but nonetheless, should not be discounted in explaining the data patterns seen this year and last. Any experienced bush teacher recognizes fluctuations from year to year in the relative slowness or brightness of his pupils. In some years the pupils seem slow and in others speed ahead. So the present possibility is certainly not unprecedented.

Of course, there are other theories to account for this data, not the least of which is the possibility of sampling error. A random sample cannot guarantee a representative cross-section of pupils selected for testing. It can only guarantee all pupils an equal chance of selection. With small samples, the probability of selecting from the low end is greater than with the larger samples. But, the probability is just as great of selecting from the top end, and with limited time and resources, these risks must be taken. While the sampling error theory is logically as sound as any other, it would be impractical to place great faith in it since it has no implications for program development. Of the two theories posed above, the one that demands consideration from a practical point of view is the first because it asks for program review by its developers and practitioners to help make sure the third level achieves a sound functioning basis.

A fourth theory, of course, is that the total concept of bilingual education is questionable and may not come through on its initial promise to provide

a quality educational program for Yupik speaking children. But such a theory is so easily weakened by the remarkable performance of children in the early primary grades experiencing the bilingual classroom.

In summary, it seems apparent that the children in the bilingual program are gaining a sound basis in nearly all aspects of Yupik literacy and oral language development. Areas of weakness noted above can, in the evaluator's opinion, be strengthened through direct attention to specific portions of the curriculum. With this in mind, the following recommendations seem in order.

PROGRAM RECOMMENDATIONS

1. The bilingual approach to math should be reviewed with a basic goal of finding some sort of compromise method of identifying numbers in the classroom. It has been suggested in numerous quarters that all phases of math use English names for numbers so that the base ten counting system can be established without linguistic confusion. Math itself, should still be taught in Yupik but with the above change. Such a change would not rule out teaching counting in Yupik in the beginning grades, but this should be treated, rightfully, as a part of language development and cultural enrichment and not a vehicle for learning math.

2. Another attempt should be made to develop math materials in the Eskimo Language Workshop. But, rather than adapt commercial textbook approaches, such as the Addison Wesly series, an effort should be made (1) to secure a math consultant to work with the Eskimo Language Workshop, and (2) develop a program of on-site implementation to help bridge the gap between newly developed math materials and their implementation in the classroom.

3. Review the alphabet concept and the role it plays in developing en-

coding skills. In reviewing the responses to this test item, the biggest problem seems to have been the intrusion of letters from the English alphabet into the thinking of the bilingually trained children. As stated before, this appears to have had little adverse effect on the bilingual porgram children since intrusions of English letters when sounds and words were encoded happened rarely. These problems are definitely more acute for the comparison school children who seem not to be able to seperate their languages cognitively, supporting the validity of the bilingual approach where this kind of confusion apparently is being circumvented.

4. It is clear that more time needs to be devoted to creative written expression in Yupik in the upper primary grades. There can be little doubt that the children in the bilingual program are gaining the basic encoding skills for writing but these skills may well be lying fallow. This problem raises the broader question concerning present limitations of bilingual teacher training posed on the bilingual teacher. Given the present eight-week preparation plus short in-service workshops, it is impressive indeed that so much has been accomplished. The time is ripe for the development of a stronger in-service training program that will free both the bilingual and the ESL teacher to become resources to one another, with the help of outside guidance from program administration and their consultants.

5. Finally, regarding the English language (ESL) portion of the curriculum, the basic objectives are being met but with some hesitancy. In the early grades (one and two) no real changes seem necessary. At grade three, creating lessons with more opportunity for the development of associative meaning could be explored, but at present there may be little more needed than to "keep the faith."

APPENDIX A

Stimulus Material for Yupik Literacy Tests

ANSWER SHEET A

SCORE _____

VISUAL DISCRIMINATION OF SYMBOLS

NAME _____

demonstration:

at atam

pi tupik tupa qupellrua

ta tema tuma atata

mu maraq murak nutek

SCORE _____

MATCHING THE WORDS WITH THE PICTURES

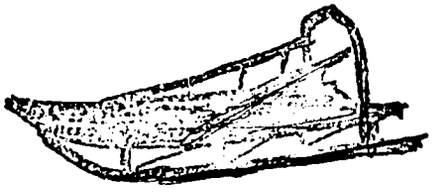
NAME _____

ena

patu

Ikamraq

saskaq



yaqulek

nutek

APPENDIX B

Instructions to Testers on Testing and Sampling

CENTER FOR NORTHERN EDUCATION

UNIVERSITY OF ALASKA
COLLEGE, ALASKA 99701

March 28, 1973

Dear

Here are some instructions for you to follow in doing the testing. By now you should have received the testing materials, or if you haven't, you should shortly.

When I got back to Fairbanks after our workshop and started working out the details the first thing I noticed was that some of you would have had a much more difficult job than others, so I went ahead and made a few changes to help even out the number of children each of you would test. I hope you won't be mad at me. I also found that the Math Test, as we originally worked it out in Bethel was too long so I shortened it a little bit and simplified it's administration (I hope).

If you look in Table I on the next page, you will see the lists of schools to be tested and who will be testing at each school. I also put down the number of children to be tested at each grade level. Following Table I is a detailed procedure for getting a random sample of children for testing. Go over it and practice it until you understand it. This is very important.

Also, when you test the children, start with Grammatic Closure and Auditory Association in Yupik. Do the rest of the tests in whatever order you think is best. Break up the testing into at least two sessions for each child, maybe even three, so he won't get too tired.

Before you go out to the other villages, you should practice all of the tests on a couple of your own pupils. To do this, you can first figure out which children in your village are going to be tested,(this will be good practice itself) and then choose a couple of children you know will not be tested to practice on. I know this all seems complicated, but if you do enough practicing before you go out, you will do a lot better job. Also, make sure the other teachers know what you are doing so they will understand too.

Good Luck,

Jim Orvik

Jim Orvik

JO/jf

TABLE I

Testing assignments and numbers of children to be tested in BIA, ASOSS and comparison schools, April 15-30, 1973.

SCHOOLS	TESTERS	NUMBER TESTED (BY GRADE)				
		K	1	2	3	TOTAL
COMPARISON SCHOOLS						
Kwigillingok	Frank Mathew	-	5	5	4	14
Eek	Sophie Parks	-	5	5	4	14
Napaskiak	Sophie Parks	-	5	5	4	14
Kwethluk	Tim Samson	-	5	5	4	14
	SUBTOTAL	0	20	20	16	56
BIA SCHOOLS						
Kipnuk	Lucy Nelson	-	5	5	-	10
Quinhagak	Lucy Nelson	-	4	4	-	8
Tuntutuliak	Lucy Nelson	-	4	4	-	8
Napakiak	Tim Samson	-	4	4	5	13
Kasigluk	Helen Nicori	-	5	5	-	10
Nunapitchuk	Zack Ivon	-	5	5	5	15
Akiachak	Helen Nicori	-	4	4	5	13
	SUBTOTAL	0	31	31	15	77
ASOSS SCHOOLS						
Bethel	Frank Mathew, Zack Ivon	10	5	5	-	20
Kongiganak	Joe Alexie	-	5	5	-	10
Twin Hills	Anecia Alakayak	-	4	3	-	7
Togiak	Anecia Alakayak	-	8	8	-	16
Manokotok	Joe Alexie	-	7	7	-	14
Aleknagik	Joe Alexie	-	4	4	-	8
	SUBTOTAL	10	33	32	-	75
	GRAND TOTAL	10	84	83	31	208

UNIVERSITY OF ALASKA

SAMPLING PROCEDURE

In order to test a random sample of children at each site, the following steps should be taken.

STEP I: Obtain an alphabetical list of boys and a list of girls for each grade level from the school's principal teacher and number the children on each list, starting at the top.

STEP 2: Remove from each list any children who have been absent from school more than ten school days in a row.

STEP 3: Determine from the schedule of testing assignments, how many children in each grade you are supposed to test at that school. (for example, Lucy is going to Kipnuk and will test 5 level one children, and 5 level two children).

STEP 4: Determine how many boys and how many girls you will test at each grade level so that the number of boys and girls is about equal. If you are supposed to test an even number, say 6, then test 3 boys and 3 girls. But if you are supposed to test an uneven number, like 5, check which list is longer and select the extra one from that list. So if there are more girls than boys and you need to test 5 children at that grade level, test 3 girls and 2 boys.

STEP 5: Use the following list of random numbers I've selected (6,3,9, 7,1,4,10,8,2,5) to determine exactly which children on each list you should test. For example, Anecia is in Togiak and she is supposed to test 8 children in grade level one. She will test 4 boys and 4 girls. Let's say that the girl's list has 8 names on it. She would then select the 6th girl on the list, then the 3rd on the list, (she can't select the 9th because there are only 8) the 7th on the list and the 1st on the list. Then she would repeat this step for the boys and girls in grade level two. This method will work for any size list up to 10. If there are more than 10 on the list, go to step 6.

STEP 6: If a particular list has more thn 10 names on it, divide the list in half and make your first selection from one list, the next selection from the other, and the next from the first list and so on. For example, let's say that one of your boys lists has 12 names on it. Divide it into two lists of 6 names each. The first child selected would be the 6th name on the first list, the second child selected would be the 3rd name on the second list and the third child selected would be the 1st name on the first list, and the fourth name on the second list. Get it? Do this until you have selected all the names you need from the list. Remember, do step 6, only if you have a list with more than 10 names on it. Otherwise you only need to do steps 1 thru 5.

Practice this in your own village with your principal teacher until you get used to it. You can impress him by explaining to him what a Random Number Table is.

UNIVERSITY OF ALASKA

Let me wish you the best of luck, and thank you for all the help you have given us so far. I look forward to hearing from you, and seeing your test results. Remember, be fair to all the children and above all be nice so they'll like you and want to do well for you.

Sincerely.

Jim Orvik

Jim Orvik

JO/jf