An Alignment/Transfer Experiment with Low Socioeconomic Level Students

By June Isaacs Elia

Background

While Lysakowski and Walhberg (1982) in their meta analysis identified three critical components of effective instruction—cues, participation, and corrective feedback—that produced approximately one standard deviation superiority over control groups, Stover (1982), Koczor (1984), Tallarico (1984), and Fahey (1986) recorded differences in test scores of over three standard deviations by simply aligning testing to instruction. Other studies have also documented successful academic achievement when there is the alignment of objectives, instruction, and assessment in the curriculum (Levine, 1982; Scott & Brock, 1983; Niedermeryer, 1979, 1977, 1976; Niedermeyer & Yelon, 1981; Niedermeyer & Sullivan, 1977; Vanderplas, Sanderson, & Vanderplas, 1964; Rosenthal & Zimmerman, 1976; Parsley, 1975; Putoff, 1973; Hagen, 1980; Ahlfors, 1979). The power of congruence between critical features of input stimuli and posttest demands has also been demonstrated (Tulving & Thompson, 1973; Baddeley, 1982).

June Isaacs Elia is an associate professor in the Department of Education at Holy Names College, Oakland, California.

In fact, 30 years ago, Carroll (1963) claimed that a fundamental component of effective instruction is the degree to which learners have a clear picture of the instructional outcome. Underlying the issue of

instructional alignment is transfer of learning, and the degree of alignment between the stimulus conditions of instruction and those of testing is a measure of transfer demand (Cohen, 1984a). When simpler tasks are included in the more complex the tasks share common elements, and transfer occurs (Resnick & Ford, 1978; Gagné, 1962; Thorndike, 1903; Snygg, 1949; Skinner, 1953).

Purpose of the Study

The present study extended the work of Stover (1982), Koczor (1984), and Talarico (1984) on instructional alignment. The study defined alignment as the behavioral congruence of the instructional process with the post-treatment tests (Cohen, 1982, 1983). The study chose to generate data about the effects of variations in the match between the stimulus conditions of instructional practice and those in testing. It measured transfer demand whenever critical features in the instructional task were not present in the post-treatment tests.

Transfer of learning was defined, on the one hand, as stimulus generalization, and, on the other hand, as response class generalization, both of which subsume the common elements construct. The purpose of the study was to determine the amount of variance explained by the alignment of testing to instruction. If traditional instruction generated .25 to .50 sigma effects (Cohen, 1987), what effect would alignment have when the critical features of the instruction were carefully matched in post-treatment tests?

Research Sample

Subjects for this study were 45 fourth-grade students randomly selected from one low socioeconomic level public school in Oakland. The school had an enrollment of 620 in grades kindergarten through sixth and received both Chapter I and EIA/LEP state and federal funds. Its multiracial students were 56 per cent African-American, 18.8 per cent Asian, 13 per cent Hispanic, 9.6 per cent caucasian, and 2.6 per cent other. Fourteen different languages were spoken. Nearly 45 per cent of its families received federal assistance through the A.F.D.C. program. There were 23 boys and 22 girls. Their ages ranged from nine to 11 years with 34 subjects between 9.0 and 9.11.

The total reading scores on the California Test of Basic Skills ranged from 1.9 to 5.5, with a mean score of 3.57 and a standard deviation of .86. Each subject was randomly assigned to one of three instructional treatment groups and received similar instruction. The instructional time for each group was systematically counterbalanced. Each subject was also randomly assigned to one of three test format groups. All subjects with outliner scores were removed from the study.

Research Design

The study used a 3x3x2 mixed experimental design—three levels of instruc-

tional practice treatment, by three levels of test item format, by two levels of word category. Instructional practice and word category were within-subjects variables, and test format was a between-subjects variable.

Treatment Procedure

All subjects were taught the meaning of 24 target words in three separate instructional treatments. Teaching materials were written in phrase, sentence, and paragraph contexts. Each instructional treatment used only one of these contexts. For example, in the phrase treatment students were taught the meaning of the target word **devoted** in a phrase context as illustrated in the following example:

Phrase Treatment
a **devoted** grandmother
a. loving
b. sick
c. weak
d. old

Sentence and paragraph treatment formats were similar as the following examples indicate:

Sentence Treatment
The sixth graders **assist** the teachers in the kindergarten room.
a. listen to
b. see
c. meet
d. aid

Paragraph Treatment
Sam kicked the ball on the roof. The portable is not **sturdy**. The children have been told many times about climbing on the roof. William climbed up to get the ball. Of course, the teacher saw William, and he was suspended from school for one week.
a. old
b. tall
c. solid
d. beautiful

Subjects were taught eight different target words in each treatment. Each target word had four correct word meanings. For example, for the target word **devoted** the correct meanings were the synonyms **faithful**, **caring**, **kindly**, **loving**. The instruction in each treatment used multiple examples of the target words in the context of the instructional treatment and followed mastery learning type modules with feedback and practice in small groups and individually until all eight target words

using all correct meanings was mastered.

The day following instruction, subjects were given the criterion tests. Each subject took two criterion tests—one for the target words, and one for the word variants of the target words. The word variants were inflected or extended forms of the target words as in the following examples:

target word	*word variant*
humor	humorous
risk	risky
yield	yielded

Fifteen subjects were randomly assigned to be tested using a phrase format, 15 using a sentence format, and 15 using a paragraph format. Only one-third of the subjects had been instructed in the same format as tested. The experiment took place on six days. Each of the three treatment days was followed by a test day. Treatments took 50 minutes; testing took approximately 10 minutes.

Hypotheses

Two instructional alignment hypotheses were proposed. Both investigated the degree of the effect in the alignment of instruction and testing of vocabulary. The first hypothesis proposed that the mean score of the aligned formats would exceed that of the misaligned. The second hypothesis proposed that the mean score of the target words would exceed the mean score of word variants of target words.

Development and Validation of the Instructional Program

The target words selected were those whose meaning would most likely be obscure to low socioeconomic level fourth-grade subjects. Their frequency level ranged from one in 100,000 running words to one in 10,000,000 (Carroll, Davies, & Richman, 1971). Reading level of the words was from grade five to 12. None of the words was listed in the school's basal reading program for grades one through four. The words were reviewed by teachers of low socioeconomic level students, and they were piloted with fifth-grade students at the same school.

In addition, fourth-grade experimental subjects were asked to write synonyms for each target word immediately before treatment. It was found that only .08 per cent of the synonyms on the test were known by the subjects before treatment. The 96 synonyms selected as correct answers for the meaning of each of the target words were words judged by classroom teachers to be within the listening-speaking vocabulary of low socioeconomic level subjects. The study was piloted three times and adjustments made after each pilot.

Analyses of Data

There were 45 subjects for each of the instructional practice conditions—

phrase, sentence, and paragraph. Fifteen subjects were tested in each of the test formats—phrase, sentence, and paragraph. Each subject's performace generated six scores, three for target words and three for word variants. Each analysis was a 3x3x2 ANOVA in which the first and last factors were repeated measures. The BMDP Statistical Software (Dixon, 1983) designed to be used with repeated measures was used to analyze the data. The analysis of variance was run on the data to analyze the amount of variance explained by various combinations of the independent variables. Appropriate testing parameters were placed. Effect sizes were calculated on all comparisons. The effect size is the difference in means of the most aligned cell with each of the other cells expressed as a percentage of the common within group standard deviation.

An additional parameter was placed on the study. A difference equal to or greater than 70 per cent of the common within-group standard deviation was defined as an educationally significant difference. This meant that there had to be a 70 per cent difference in scores for the study to be considered of significance to classroom teachers.

Findings

The first hypothesis was confirmed; the pooled means of the aligned was 7.20 with a standard deviation of .73. The pooled aligned means exceeded the pooled means of the misaligned by 91 per cent of the common within-group standard deviation. This meant that subjects performed 91 per cent better when they were tested using the same format as in instruction. (The aligned means for each treatment are underlined in Table 1.)

The second hypothesis was also confirmed. The mean score of target words exceeded the mean score of word variants by 85 per cent of the common within-group standard deviation. Student performance on the test of target words was over three-quarters of a standard deviation above that of scores on variants of the target words.

Of the six cell comparisons for target words, all were in the direction hypothesized with four generating effect sizes exceeding 70 per cent of the common within-group standard deviation. The analysis of pairwise comparisons showed that the mean of the target words exceeded the mean of the word variants in eight of the paired comparisons. Eight of the nine word variant comparisons generated effect sizes exceeding 70 per cent of the common within-group standard deviation. Viewing the study as a whole out of the total 15 possible pairwise comparisons, 12 were consistent with the research prediction.

The ANOVA testing the first hypothesis indicated that alignment/misalignment alone accounted for 16 per cent of the total variance; and test format accounted for seven per cent. The study explained 23 per cent of the variance. The ANOVA testing the second hypothesis showed that word category alone accounted for 18 per

Table 1
Means and Standard Deviations of Target Words and Word Variants by Treatment and Test Format

	Treatment					
	Phrase		Sentence		Paragraph	
	Mean	SD	Mean	SD	Mean	SD
Test Format						
Phrase						
target word	7.80	.40	6.80	1.22	5.47	1.02
word variant	5.20	1.97	5.20	2.10	4.93	1.24
Sentence						
target word	7.33	.87	7.47	.72	5.73	1.12
word variant	6.53	1.41	6.67	1.19	5.93	1.39
Paragraph						
target word	6.27	1.34	6.33	1.58	6.33	1.08
word variant	4.67	1.35	5.60	1.40	5.27	1.53

The underlined means are the values for the aligned condition under each treatment that were used to estimate a pooled mean.
The pooled mean for the aligned cells was 7.20 with a standard deviation of .73; the pooled mean of misaligned cells was 6.32 with a standard deviation of 1.19.

Table 2
Magnitude of Effect Sizes Among Treatment, Test Format, and Word Category

	Testing	
	Target Words	Word Variants
Phase Treatment		
Phrase Format		2.19*
Sentence Format	.74*	1.40*
Paragraph Format	1.76*	3.58*
Sentence Treatment		
Phrase Format	.69	1.61*
Sentence Format		.84*
Paragraph Format	.99*	1.95*
Paragraph Treatment		
Phrase Format	.82*	1.21*
Sentence Format	.55	.33
Paragraph Format		.82*

Parameters for Hypothesis 1 were: a=.05, b=.05, d=.70, n=45.
Parameters for Hypothesis 2 were: a=.05, b=.05, d=.70, n=135.
* Effect sizes equal to or greater than the predicted d = .70.

Table 3
Analysis of Variance with Repeated Measures: Data for Hypothesis 1

Source	Sum of Squares	df	Mean Square	F	p	Explained Variance
Between Subjects	93.97	44				
Test Format	6.77	2	3.39	1.63	.20	.07
Error	87.20	42	2.08			
Within Subjects	142.00	90				
Main Effect[1]	23.12	1	23.11	33.48	.00	.16
Main Effect[2]	13.39	2	6.69	9.69	.00	.09
Error	29.00	42	.69			
Misalignment[3]	20.54	1	20.54	19.64	.00	.15
Misalignment[4]	12.02	2	6.01	5.75	.01	.09
Error	43.93	42	1.04			

[1]Alignment vs. Misalignment; [2]Alignment/Misalignment x Test Format; [3]Misalignment 1 vs. Misalignment 2; [4]Misalignment 1/Misalignment 2 x Test Format.
Test Parameters: a = .05 b = .05 d = .70 n = 45

Table 4
Analysis of Variance with Repeated Measures Data for Word Category: Hypothesis 2

Source	Sum of Squares	df	Mean Square	F	p	Explained Variance
Between Subjects	251.19	44				
Test Format	40.54	2	20.27	4.04	.02	.16
Error	210.64	42	5.02			
Within Subjects	439.33	225				
Category 1	78.95	1	78.95	45.62	.00	.18
Category 2	14.36	2	7.18	4.15	.02	.03
Error	72.69	42	1.73			
Treatment 1	28.94	2	14.47	9.26	.00	.07
Treatment 2	27.10	4	6.78	4.34	.00	.06
Error	131.29	84	1.56			
Treatment 3	16.23	2	8.11	10.80	.00	.04
Treatment 4	6.66	4	1.66	2.22	.07	.02
Error	63.11	84	.75			

Category 1 - Word Category; Category 2 - Word Category x Test Format; Treatment 1 - Instructional Treatment; Treatment 2 - Instructional Treatment x Test Format; Treatment 3 - Word Category x Instructional Treatment; Treatment 4 - Word Category x Instructional Treatment x Test Format.
Test Parameters: a = .05 b = .05 d = .70 n = 135

cent of the total variance; test format 16 per cent; and treatment seven per cent. The total explained variance was 41 per cent.

All second-order interactions for treatment and test format were significant. The phrase testing format generated both the highest and lowest scores. The mean score under the paragraph format remained constant across treatments. Six mean scores on the word variants fell below the lowest target word mean score. The word variant sentence format generated the highest mean score. The paragraph format means were lowest on the phrase treatment, rose slightly above its target word mean on the sentence treatment, making the second highest score, and dropped to its second place position on the paragraph treatment. The mean score under phrase format were almost even across all treatments, slightly lower on the paragraph treatment.

Discussion and Conclusions

The most significant implication of this study for teacher educators rests in the power of the instructional alignment construct with the low socioeconomic level students in this study. Alignment did have an effect, an unusually high effect. This study found that a change in the test format from conditions of instruction accounted for a 91 per cent correct difference in score while a change in word category accounted for an 85 per cent correct difference in score. A continuum of achievement from the most aligned to the greatest transfer demand was measured to be as much as a 358 per cent correct difference in score. This difference of from one to three and one-half standard deviations is considerable, especially for the low socioeconomic level low achiever. If traditional instruction generated only .25 to .50 sigma (Cohen, 1987; Walberg, 1990), and alignment generated from .33 to 3.58 sigma, alignment must be considered a critical component of effective instruction. The alignment effect for group instruction nearly matched Bloom's (1984) "four-to-one" effect for tutorial instruction. Recent studies continue to corroborate the effect of instructional alignment (Nolen, 1991; Boris, 1991; Davis, 1991).

High socioeconomic level high achievers scored 41 per cent better on arithmetic tests aligned to instruction (Koczor, 1984), while the present study with low socioeconomic level underachievers scored 91 per cent higher on aligned tests. This difference in score meant that a student who scored at the 50th percentile would score at the 80th percentile simply by aligning the test to the instruction. The alignment effect for low socioeconomic level low achievers appears to have washed out achievement and supports Bloom's propostion (1976) that given appropriate instruction, low aptitude students can be made to perform as well as high aptitude students. Classroom teachers may not consider it very different to test the phrase "sense of **humor**" by using any of the following formats:

"a **humorous** story" (Effect: 2.19 sigma)
"My mother has a good sense of **humor**." (Effect: .74 sigma)

"I read a **humorous** story about a monkey". (Effect: 1.40 sigma)
"I can predict that my teacher will laugh at your joke. He has a great sense of **humor**..." (Effect: 1.76 sigma)
"It wasn't at all **humorous** to watch the Raiders leave Oakland. Our family never missed a game..."(Effect: 3.58 sigma)

However, these slight misalignments made great differences in test scores. For example to test **a humorous story** instead of the phrase **sense of humor** which was taught made a 219 per cent difference in test score. Testing the sentence **I read a humorous story about a monkey** produced a 140 per cent difference in test score. The effect sizes ranged from .74 to 3.58 sigma for the various testing formats of the target word **humor** and the word variant **humorous**.

A very practical implication for teacher education on lesson and curriculum planning would be to clearly define the instructional outcomes, and to carefully arrange the stimulus conditions of instruction and testing to match those desired outcomes. This would appear to be especially warranted practice when teaching the low socioeconomic level low achiever. To do otherwise would mean that educators might simply be measuring what was clearly not taught.

> It appears that whatever else scholastic aptitude means, it manifests itself in the school setting by the degree to which learners depend upon instruction to perform on school tests. Apparently, when those tests demand responses varying from instruction, some children have in their repertoires behaviors learned outside the classroom or beyond the scope of formal instruction that carry them through successfully. We say these children can transfer learning, and we furthermore call them "smart," "high aptitude," and "high ability." They seem to learn no matter how little they are taught because they depend less upon the lesson than children who know less. To **know less** may often mean that under varying stimulus conditions, they do not "transfer". (Cohen, Busk, & Koczor, 1984, p.7)

The low socioeconomic level low achievers in this study demonstrated lower and lower test scores on successive and continued tests of misalignment. This study demonstrated that low performing low socioeconomic level students could achieve to high levels of success when conditions of testing matched those of instruction. It does not suggest to educators **what** to teach, but rather **how** (Cohen, 1987; Clark, Blake, & Knostman, 1989; Cohen, Hyman, Ashcroft, & Loveless, 1989). It suggests a fine tuning of teaching and testing for the low socioeconomic level low achievers.

References

Ahlfors, G. (1980, May). Learning word meanings: A comparison of three instructional procedures. Doctoral disseration, University of Minnesota, 1979. *Dissertation Abstracts International*, 40, (11). University Microfilms No. 8011754.

Baddeley, A.D. (1982). Domains of recolletion. *Psychological Review*, 89, 708-729.

Bloom, B.S. (1976). *Human characteristics and school learning*. New York: McGraw Hill.

Bloom, B.S. (1984, June/July). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13, (6), 4-16.

Boris, V. (1991). Using controlled degrees of misalignment to teach cognitive stretching: A set for transfer of learning mathematics and vocabulary. Doctoral dissertation, University of San Francisco.

Carroll, J.B. (1963). A model of school learning. *Teachers College Record*, 64, 723-733.

Carroll, J.B., Davies, P., & Richman, B. (1971). *Word frequency book*. Boston, MA: Hougton Mifflin Company.

Clark, R.E., Blake, S., Knostman, V. (1989). Metacognitive processes supporting far transfer: Selecting and connecting. Paper presented to the American Educational Research Association, San Francisco, CA.

Cohen, S.A. (1982). Two components of quality instruction: alignment and P ratio. Paper presented at the Annual Meeting of the American Educational Research Association, New York.

Cohen, S.A. (1983). Instructional psychological implications of mastery learning. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada, Session 5.09.

Cohen, S.A. (1984a). A rationale for alignment. A paper delivered at the University of San Francisco.

Cohen, S.A. (1984b). Implications of instructional psychological research on mastery learning. *Outcomes*, A Quarterly Newsletter of the Network of Outcome-Based Schools, 2, 18-25.

Cohen, S.A. (1987, Nov.) Instructional alignment: Searching for a magic bullet. *Educational Researcher*, 16-20.

Cohen, S.A., Hyman, J.S., Ashcroft, L., & Loveless, D. (1989). Comparing effects of metacognition, learning styles, and human attributes with alignment. A paper presented at the American Educational Research Association, San Francisco, CA.

Cohen, S.A., Busk, P., & Koczor, M.L. (1984). Experiments in the effects of instructional alignment. A paper delivered at the University of San Francisco.

Davis, J.M. (1991). Effects of cognitive entry level behaviors on the reading performances of limited English speaking students. Doctoral dissertation, University of San Francisco.

Dixon, W.J. (Ed.) (1983). *Statistical Software 1985 Reprinting*. Berkeley, CA: University of California Press.

Fahey, P.A. (1986). Learning transfer in main idea: Effects of instructional alignment and aptitude on student scores on main idea test item. Unpublished doctoral dissertation, University of San Francisco.

Gagne', R.M. (1962). The acquisition of knowledge. *Psychological Review*, 69, (4), 355-365.

Hagen, J.E. (1980, June). The effects of selected prereading vocabulary building activities on literal comprehension, vocabulary understanding and attitude of fourth and fifth grade students with reading problems. Doctoral dissertation, The University of Wisconsin, 1979. *Dissertation Abstracts International*, 40, (12-A), 6216.

Koczor, M.L. (1984). Effects of varying degrees of instructional alignment in posttreatment tests on mastery-learning tasks of fourth-grade children. Unpublished doctoral disser-

tation, University of San Francisco.

Levine, D.U. (1982, April). Successful approaches for improving improving academic achievement in inner-city elementary schools. *Phi Delta Kappan*, 523-526.

Lysakowski, R.S., & Walberg, H.J. (1982, Winter). Instructional cues, participation, and corrective feedback: a qualitative synthesis. *American Educational Research Journal*, 19, (4), 559-578.

Niedermeyer, F.C. (1976, Nov.) Developing classroom based instructional products: An evolving set of guidelines. *Educational Technology*, 55-57.

Niedermeyer, F.C. (1977, January). A basis for improved instructional leadeship. *Elementary School Journal*, 248-254.

Niedermeyer, F.C. (1979, Aug.6). *Curriculum alignment—A way to make schooling more understandable*. Professional Paper 41, Los Alamitos, CA: Southwest Regional Laboratory for Educational Research and Development.

Niedermeyer, F.C., & Sullivan, H.J. (1977, June). Prospects for school acceptance of objectives-based instructional programs. *Educational Technology*, 1-25.

Niedermeyer, F.C., & Yelon, S. (1981, May). Los Angeles aligns instruction with essential skills. *Educational Leadership*, 38, (8), 618-620.

Nolan, L. (1991). Cognitive stretching: Teaching college students a set for transfer using an instructional alignment model. Doctoral dissertation, University of San Francisco.

Parsley, N.J.J. (1975, Oct.). Rule learning and transfer in oddity and discrimination problems. Doctoral dissertation, Tulane University, 1975. *Dissertation Abstracts International*, 36, (4-B), 1951.

Putoff, O.E. (1973, March). An evaluation of transfer following discovery oriented tutoring of young children on classification and related tasks. Doctoral dissertation, Clarement Graduate School, 19973. *Dissertation Abstracts International*, 33, (9-B), 4494.

Resnick, L.B. (1976). Task analysis in instructional design: Some cases from mathematics. In D. Klahr (Ed.), *Cognition and instruction*. Hillsdale, NJ: Erlbaum.

Resnick, L.B., & Ford, W. (1978). The analysis of tasks for instruction: An information-processing approach. In A.C. Catania, & T.A. Brinham (Eds.) *Handbook of Applied Behavior Analysis*, (pp. 378-409). New York: Irvington Publishers, Inc.

Resnick, L.B., Wang, M.C., & Kaplan, J. (1973). Task analysis in curriculum design: A hierarchically sequenced introductory mathematics curriculum. *Journal of Applied Behavior Analysis*, 6, 679-710.

Rosenthal, T.L., & Zimmerman, B.J. (1976). Organization and stability of transfer in various concept attainment. *Child Development*, 47, 110-117.

Scott, R.O., & Brock, W. (1983, Winter). *Curriculum alignment: Improving instruction in essential skills. File three: tests, texts*. Los Alamitos, CA: Southwest Regional Laboratory for Educational Research and Development.

Skinner, B.F. (1953). *Science and human behavior*. New York: The Free Press.

Snygg, D., & Combs, A.W. (1949). *Individual behavior*. New York: Harper & Brother Pub.

Stover, G. (1982). Structural variables affecting word problem solving difficulty in sixth graders. Doctoral dissertation, University of San Francisco, 1979. *Dissertation Abstracts International*, 42, (12-A), 5050.

Tallarico, I. (1985). Effects of ecological factors on elementary school student performance on norm-referenced standardized tests: Non reading behaviors. Doctoral dissertation, University of San Francisco, 1984. *Dissertation Abstracts International*, 45, (12-A),

3623.

Thorndike, E.L. (1903). *Educational psychology*. New York: Lemcke & Buechner.

Tulving, E. & Thompson, D.M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80, 352-373.

Vanderplas, J.M., Sanderson, W.A., & Vanderplas, J.N. (1964). Some task-related determinants of transfer in perceptual learning. *Perceptual and Motor Skills*, 18, 71-80.

Walhberg, H. (1990). Productive teaching and instruction: Assessing the knowledge base. *Phi Delta Kappan*, 71, 470-478.