DOCUMENT RESUME

ED 142 560 TM 006 072

AUTHOR Bond, James T.
TITLE The High/Scope Productive Language Assessment Tasks: Effects of the Cognitively Oriented Curriculum on Follow Through Children's Written Language Production.
INSTITUTION High/Scope Educational Research Foundation, Ypsilanti, Mich.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
PUB DATE 75
CONTRACT OEG-0-70-4919
NOTE 44p.; Appendices--examiner's manual and scoring manual--are copyrighted and therefore not available from EDRS. They are not included in the pagination
AVAILABLE FROM High/Scope Educational Research Foundation, 600 North River St., Ypsilanti, Michigan 48197 (with appendices, $3.00)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Cognitive Processes; Composition Skills (Literary); Evaluation Criteria; Evaluation Methods; Grade 2; Grade 3; Instructional Innovation; Language Experience Approach; *Language Instruction; Language Tests; Measurement Techniques; *Performance Tests; Primary Education; *Test Construction; Test Reliability; Test Validity; *Writing Skills
IDENTIFIERS *High Scope Cognitively Oriented Curriculum; Project Follow Through

ABSTRACT

The Cognitive Curriculum uses a nontraditional approach to language instruction. Children play an active role, learning from concrete experiences while writing and reading their own material. This type of instruction has been developed with Project Follow Through students and has been particularly successful in writing instruction. The Productive Language Assessment Tasks (PLAT) were designed to assess the impact of this program. This task-oriented instrument is suitable for measuring nontraditional objectives, and serves as an alternative to standardized tests. The PLAT was administered to second and third grade children enrolled in the Follow Through program in five different U.S. locations. A comparable non-Follow Through sample was located in four of the five sites, and given the same test. Writing samples were produced as a result of the test, and reviewed for a number of variables. These include fluency; syntactic maturity; vocabulary diversity; descriptive quantity; diversity and scope; reporting quality; narrative organization; explanatory statements; and decodability. Results indicated that by the third grade, positive effects of the Follow Through program were evident. Reports and narratives written by Follow Through children were more complex and more fluent than those written by non-Follow Through children. (GDC)

THE HIGH/SCOPE PRODUCTIVE LANGUAGE ASSESSMENT TASKS: EFFECTS OF THE COGNITIVELY ORIENTED CURRICULUM ON FOLLOW THROUGH CHILDREN'S WRITTEN LANGUAGE PRODUCTION

Fall 1975

Prepared by

James T. Bond

High/Scope Educational Research Foundation
600 North River Street
Ypsilanti, Michigan 48197

The research reported herein was performed pursuant to Grant No. OEG-0-70-4919 with the Office of Education, U.S. Department of Health, Education and Welfare. Contractors undertaking such projects under government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Point of views and opinions stated do not, therefore, necessarily represent official Office of Education opinion or policy.

Acknowledgments

Many persons have contributed to the development of the Productive Language Assessment Tasks. Betty Couvares, John Love, David Weikart, Charles Hohmann, Sheila Mainwaring, Mary Morris, and Jan Kittel have played important roles in the development of both the writing tasks and the language analysis procedures described here. Through Follow Through sponsors' collaborative efforts to develop productive language measures, persons from other Follow Through sponsors have also made important contributions to High/Scope's development efforts--particularly Carol Weissman (Education Development Center) and Allan Abrams (Far West Educational Laboratory).

The collection of data would not have been possible without the help of Mary Morris and Stewart Gordon in training testers and of project directors, curriculum assistants, teachers, and principals in permitting and coordinating data collection efforts on site.

The processing of written language samples was coordinated by Jan Kittel with assistance from Nancy Naylor and Bob Hanvey. Jan Kittel, Jane Oden, Barb Breummer, Jodi Breummer, and Sheila Roberts scored all PLAT stories. Bob Hanvey developed the computer program which was used to facilitate scoring of 1974-75 PLAT data.

Mel Shelly performed the bulk of the data analysis for this study. Judy McNeil contributed sections dealing with the effects of task order and interval between tasks on PLAT performance. John Love and David Weikart made invaluable suggestions concerning the interpretation and presentation of findings.

The text of this chapter was typed by Cathy Dice and Jan Grimston; tables were drafted by Jan Kittel and typed by Lynne Dermody.

INTRODUCTION

Unlike traditional approaches to education which are preoccupied with the content and form of children's responses to standard and convergent stimuli, the Cognitive Curriculum is primarily concerned with the processes of thinking and behaving in unstructured situations. Children assume active roles in the educational process and are encouraged to apply their intellects to make their own sense out of the world around them. Concrete experiences, rather than language experiences (i.e., being told), provide the foundation of learning. Children's experiences in Cognitively Oriented classrooms are designed to foster the development of their abilities to initiate independent learning activity, to define and solve practical problems, to work cooperatively toward common goals, and to express their own ideas as well as obtain information from other sources. Such outcomes cannot be evaluated by simply asking children to dredge up from memory a particular fact, rule or algorithm--that is, by administering a standardized achievement test.

Consequently, High/Scope undertook the development of alternative procedures for assessing the impact of its educational program on elementary school children. These efforts began with High/Scope's participation in Project Follow Through in 1968 and have continued to the present. Developmental efforts have been particularly successful in the area of children's written language production.

In the Cognitive Curriculum, children do not receive programmed instruction or drill in language arts. Children are introduced to written language by first seeing their own spoken words written down by teachers, then writing about their own experiences and ideas. Reading skills begin to develop as children read their own language written down by teachers, then by themselves. A central objective of this approach is to foster strong interests

in writing and reading as useful forms of communication. The child's interest then provides a motivational basis for learning the conventions of written language in order to communicate more effectively. Children are helped to resolve particular problems in phonics as they confront these problems while trying to write what they want to write and to read what they want to read. Teachers help children learn the rules of standard punctuation when the children themselves want to punctuate their written statements in order to resolve ambiguities perceived by readers. As their reading skills improve, children are encouraged to read materials written by others in order to obtain information they seek or simply for pleasure. Each child is primarily responsible for selecting reading materials appropriate to his abilities and interests. Children's vocabularies (oral and written) reflect their individual interests and concrete experiences rather than the content of basal readers. In short, children acquire the conventions of written language in a systematic way as these conventions assume functional utility in the context of their own learning activities and the social environment of the classroom, not when achievement test norms dictate.

In order to tap important child outcomes in the area of language arts, High/Scope developed the Productive Language Assessment Tasks (PLAT). Unlike achievement and aptitude tests, the PLAT allow children to work with real objects, structuring and solving problems of their own design. Further, social interaction and cooperation among children is encouraged during all phases of the tasks. Finally, the tasks elicit written representations founded in immediate and concrete experiences and structured largely by the child rather than by stimuli associated with the test.

A preliminary version of the PLAT was field tested during the 1972-73 school year (Couvares and Love, 1973). In 1973-74, a revised version of the instrument was administered to more than 200 Follow Through and non-Follow Through children at four of High/Scope's Follow Through sites. The results of this study and detailed descriptions of administration and scoring procedures can be found in an addendum to High/Scope's 1973-74 Follow Through Report (Love and Bond, 1975). The PLAT battery was further revised for use in 1974-75 Follow Through evaluation activities. The research design and methodology and the major findings from 1974-75 evaluation research are presented in this chapter.

RESEARCH DESIGN AND METHODOLOGY

Sampling

In the spring of 1975, the High/Scope Productive Language Assessment Tasks (PLAT) were administered to virtually all second and third grade children who had been enrolled in the Follow Through program since entering school at five of the sponsor's centers: Leflore County, Mississippi; Okaloosa County, Florida; P.S. 92, New York City; Howland and Lathrop Schools, Chicago; and Central Ozarks, Missouri. The PLAT were also administered to groups of non-Follow Through children at four of five sites. At one site--New York City--it was not possible to secure a comparison group. On site project staff were responsible for the selection of non-Follow Through children. Although procedures varied somewhat from site to site, an attempt was made to roughly match non-Follow Through with Follow Through children on grade level, sex, ethnicity, economic status, and residential mobility.

Validity of Follow Through/non-Follow Through Comparisons

The adequacy of non-Follow Through comparison groups cannot be directly evaluated. Available evidence, however, suggests that comparisons may be biased in favor of non-Follow Through samples. Historically, Mississippi has provided the most convincing Follow Through/non-Follow Through comparisons. There is a large, fairly homeogeneous Follow Through eligible population in Leflore County. Comparison groups have always been drawn from potentially eligible children attending a local non-Follow Through school. Even in Mississippi, however, spring 1975 interviews with a subsample of third grade Follow Through and non-Follow Through parents indicated that mothers of non-Follow Through children had more formal education than mothers of non-Follow Through children. This was also true in the Missouri sample. In addition, the Missouri comparison group was drawn from the same schools attended by Follow Through children. Consequently, a Follow Through diffusion effect may be operating there. Moreover, there is some reason to believe that assignment to Follow Through or non-Follow Through classrooms is not random among eligible children. In Florida, it has always been necessary to scour the countryside to form a comparison group, and the adequacy of these groups has always been suspect. A survey of Florida Follow Through children conducted in 1974 by school

psychologists strongly suggested that Follow Through classrooms had become repositories for special education eligible children who could not be integrated into other classrooms. In Chicago, the 1975 comparison group was selected by the principal of a non-Follow Through school. There are indications that selection criteria were applied loosely and to the advantage of the comparison group.

Thus, it is with qualification that the findings of this research are interpreted. Although the results of Follow Through/non-Follow Through comparisons generally favor the Follow Through group, it seems likely that the observed differences would be of even larger magnitude if more adequate comparison groups were available.

Measurement

All data were collected in spring 1975. The PLAT were administerd to all children in the sample. PLAT testers were hired at each site and trained by High/Scope staff. All testers had previous experience working with children of second and third grade age. The instrument is described below.

In addition to the PLAT, standardized aptitude and achievement test data were obtained on some children. For the most part these data were obtained in local school district testing. However, Stanford-Binet Intelligence Scales were administered by psychologists under contract with High/Scope at two sites. The purpose of collecting aptitude and achievement data was to further the development of the PLAT rather than to evaluate program effects. All of these measures are well known and will not be described here.

Productive Language Assessment Tasks

The PLAT battery incorporates two tasks--Reporting and Narrating. Detailed descriptions of both tasks appear in Appendix A to this volume. The basic materials and administrative procedures for each task are briefly summarized here.

Reporting task. In the Reporting task, children were given identical sets of unstructured materials and were asked to make anything they wanted to make. Materials included: paper of different grades, colors, and shapes; plastic foam, crêpe paper; double-knit fabric; plastic screen; paper fasteners; cotton; pipe cleaners; and a rubber band. Children

could also use scissors, magic markers, and tape available in the examination room. Children were allowed 20 minutes to make something and were permitted to keep whatever they made along with any materials that remained. After 20 minutes, children were asked to write about "how they made" whatever they made and were allowed 30 minutes to complete their stories. Children were permitted to interact with one another during all phases of the task.

Narrating task. In the Narrating task, each child was given a set of relatively unstructured materials to "help you make up a story". Materials included: 12 one-inch cubes of two colors; a wooden "car" with bottle cap wheels; abstract human figures made from wooden dowels of different lengths; 2 pieces of felt; and a cardboard box. After about 15 minutes of free (and usually dramatic) play on a carpeted floor, children were asked to write a "make-believe or pretend story". As in the Reporting task, children were permitted to interact with one another as they played and wrote.

Variables. Initial processing of writing samples produces thirty-two major first-order variables--16 for each task. Eleven second-order variables--those analyzed in this report and described below--are constructed from the first-order variable set during subsequent computer processing. All but two of these eleven variables are derived by combining Reporting and Narrating scores. The purpose for combining the two writing samples is to obtain a more representative corpus of individual written language production and, thus, more reliable individual measures. Specific criteria for scoring are described in Appendix B to this volume.

1. Fluency. Fluency is an indicator of writing facility apart from any consideration of writing quality. Computationally fluency equals the mean number of words in the valid texts of Reporting and Narrating stories.

2. Syntactic Maturity. The syntax of oral and written language undergoes fairly systematic changes during childhood. Syntactic maturity is estimated by measuring the average length of T-units in children's stories.

   A T-unit (Hunt, 1965) is a single independent predication (subject + verb + object if verb is transitive) together with any subordinate clauses or phrases that

may be grammatically related to it. It may be a simple or complex sentence but not a compound sentence. In conversation and written dialogue, elliptical constructions are accepted as T-units if missing grammatical elements are clearly implied by preceding T-units. The number of T-units in a language sample represents the number of grammatically complete statements.

Average T-unit length is determined by dividing the total number of words in a language sample by the total number of T-units. Research by Hunt (1965) and O'Donnell, Griffin, and Norris (1967) indicates that average length of T-unit is a valid index of syntactic maturity in both oral and written language production.

3. Vocabulary Diversity. Vocabulary Diversity is a proportional measure of diversity in the vocabulary of a language sample adjusted for length of sample (cf., Type/Token Ratio in Carroll, 1964). Diversity is computed first for each task then averaged across tasks:

$$\text{Diversity} = \frac{\text{number of different words}}{\sqrt{2(\text{number of decodable words})}}$$

4. Descriptive Quantity. Descriptive Quantity represents the total number of words and larger constructions which describe the attributes of and relationships between objects, persons, and events. It is a measure of content rather than form. The component variables are:

   classification words
   subjective modifiers (Narrating task only)
   space words
   time words
   seriation words
   physical quantity words
   number words
   transformation-combination verbs (Reporting task only)
   expressions of class relationship
   occurrences of simile and metaphor

Descriptive Quantity is computed first for each task then averaged across tasks.

5. Descriptive Density. Descriptive Density is a measure of descriptive words (defined above) as a proportion of all decodable words in the text. Expressions of class relationship and similes/metaphors are excluded since they are not expressed through single words.

6. Descriptive Diversity. Descriptive Diversity is a proportional measure of the diversity of descriptive words (excluding class relationship and simile/metaphor) adjusted for the total number of descriptive words in a language sample:

$$= \frac{\text{number of different descriptive words}}{\sqrt{2}(\text{number of descriptive words})}$$

Descriptive Diversity is first computed for each task then averaged across tasks.

7. Descriptive Scope. Descriptive Scope represents the average number of descriptive categories (see #4 above) used in Reporting and Narrating stories, disregarding how often each was used. It indicates the conceptual breadth of descriptions in reports and narratives.

8. Reporting Quality. Reporting Quality is derived from analysis of Reporting stories and represents the degree to which a report describes "how" something was made.

   1 = report is irrelevant to task
   2 = report merely enumerates materials used
   3 = report describes what was made but not how
   4 = report describes how something was made

9. Narrative Organization. Narrative Organization is derived from analysis of Narrating stories and measures the organizational quality of a narrative.

   1 = T-units in narrative are unrelated
   2 = T-units in narrative are related to one another logically and thematically, but there is no closure to the story (i.e., it might go on indefinitely)
   3 = T-units are interrelated and the narrative has closure

10. Explanatory Statements. This score represents the average number of statements in reports and narratives which express cause, rationale, and purpose to explain relationships, attributes, decisions, and events.

11. Decodability. Decodability is a measure of the degree to which a story can be decoded by a reader (in this case, scorers who are experienced readers of children's writing):

$$= \frac{\text{number of decodable words in valid T-units}}{\text{total number of words in language sample}}$$

The denominator includes all decodable words in valid T-units, all nonsense words in valid T-units, and all words in extraneous material (defined in Appendix B).

All textual analysis was done by trained High/Scope Foundation staff. Text was first edited then entered into a computer program at remote terminals in T-unit segments. Scorers coded the edited text as it played back on the terminal. The computer program then tallied all codes and computed all variables except Reporting Quality, Narrative Organization, and Explanatory Statements.

METHODOLOGICAL FINDINGS

Scoring Reliability

For the purpose of calculating interscorer agreement, fifteen samples of writing from the Reporting task and fifteen from the Narrating task were randomly selected and coded by all scorers. Since the coding of these stories occurred throughout the coding period, estimates of reliability apply to all data reported here.

Ebel's intraclass correlation coefficient (Guilford, 1954, p. 395) was used to estimate scoring reliability. It is analogous to an average intercorrelation for all possible pairs (10 in this instance) of k (5) raters. Interscorer reliability coefficients are reported in Table 1 for first-order variables coded for Reporting and Narrating stories. The major PLAT variables analyzed in this report were derived from these scores according to computational procedures described above. Scoring reliability was high; all coefficients exceed .90 except for Number of Words in Extraneous Material when scored for Reporting stories where the coefficient is .89. These findings suggest that all first-order variables are fully operationalized at this point in the development of the instrument.

Instrument Reliability

Given high interscorer reliability there seems little doubt that the PLAT is a reliable group measure, i.e., that mean PLAT scores for fairly large groups of children are replicable. The PLAT's reliability as a measure of an individual child's language, however, still needs to be determined. A small scale study is planned for 1975-76 to directly assess the reliability of the PLAT as an individual measure by establishing test-retest reliability.

Reasoning backwards from findings more pertinent to the assessment of instrument validity, there are indications that the PLAT battery provides reliable individual measures. When Follow Through children's performance at second grade is correlated with their performance at third grade, some indication of stability in individual performance is found (Table 2). Table 2 reports second (1974) with third (1975) grade correlations of four PLAT variables

Table 1

Inter-Scorer Reliability Coefficients
Computed for First-Order Variables
from which Major PLAT Variables Were Computed

Variables	Reporting Task	Narrating Task
Number of Words	.993	.999
Number of Decodable Words	.993	.999
Number of T-Units	.998	.999
Number of Different Decodable Words	.995	.996
Classification Words	.999	.973
Subjective Modifiers	NA	.985
Class Relationship	1.000	.986
Space Words	.974	.979
Seriation Words	*	.974
Physical Quantity Words	*	.978
Number Words	.999	.991
Time Words	.989	.970
Transformation-Combination Verbs	.993	NA
Reporting Quality	.935	NA
Narrative Organization	NA	.908
Number of Words in Extraneous Material	.892	.939
Number of Similes and Metaphors	*	.972
Number of Explanatory Statements	*	.940

NA = Not applicable to task

* = No occurrences in reliability sample

Table 2

Correlation of Second (1974) with Third (1975) Grade PLAT Scores for Follow Through Children in the Longitudinal Sample

Variables	Mississippi *(N=32)*	Missouri *(N=21)*	Chicago *(N=7)*
Fluency	.32	.62	.73
Syntactic Maturity	.04	.43	-.19
Descriptive Quantity	.28	.75	.93
Descriptive Density	.26	.32	.89

for three small longitudinal samples. Fluency and Syntactic Maturity were scored in the same way in both years. Descriptive Quantity and Descriptive Density were scored in very similar, but not identical, manners in both years. Other variables were not comparable across years, thus, were not analyzed. Fluency and Descriptive Quantity seem quite stable from second to third grades in both Missouri and Chicago. Syntactic Maturity evidences moderate stability in Missouri. Descriptive Scope is highly stable in Chicago. Given the length of the interval between tests, the magnitude of correlations is surprising. This is particularly so when one considers that (1) children are just beginning to acquire writing facility at second grade and (2) a rather uncommon educational treatment (particularly in the area of language arts) occurred during the interval.

Correlations between children's performance in the Reporting and Narrating tasks also suggest instrument reliability. The Reporting and Narrating tasks are not alternate forms of the same test; rather, they were designed to tap presumably different dimensions of written language production. Consequently, it was not expected that performance on the two tests would be the same. However, it was anticipated that some aspects of written language production might generalize across situations, assuming that a sufficiently representative language sample was obtained within each task. Correlations between Reporting and Narrating task variables are reported in Table 3 for the total sample and for second and third grades. Moderate to strong correlations were found for Fluency, Vocabulary, Diversity, Descriptive Quantity, Descriptive Diversity, and Descriptive Scope. These findings are at least compatible with high instrument reliability.

Finally, findings reported in the next section (Instrument Validity), indicating consistently higher performance by third, compared with second, graders and correlations of PLAT scores with aptitude and achievement test scores also suggest that the PLAT battery behaves like a reliable instrument. Final confirmation of instrument reliability will depend on the results of further developmental research.

Instrument Validity

The PLAT has high content validity in that it samples the kind of language production encouraged by the High/Scope Cognitively Oriented Curriculum and analyzes the language

Table 3

Correlation of PLAT Reporting Task with Narrating Task Variables

PLAT Variables	TOTAL SAMPLE *(N=883)*	Second Grade *(N=446)*	Third Grade *(N=437)*
Fluency	.63	.62	.61
Syntactic Maturity	.20	.24	.15
Vocabulary Diversity	.49	.44	.46
Descriptive Quantity	.52	.47	.51
Descriptive Density	.22	.19	.23
Descriptive Diversity	.45	.34	.49
Descriptive Scope	.45	.35	.47
Explanatory Statements	.23	.17	.24
Decodability	.32	.27	.38

which children produce in curriculum-relevant ways. Moreover, responses from educators not associated with the High/Scope program suggest that the PLAT battery has substantial, and fairly general face validity, i.e., it is an appropriate measure of general educational goals and real-world competencies.

Some findings related to the construct validity of the instrument can also be adduced. Written language production is presumed to "improve" (increase in amount, syntactic quality, semantic complexity, and logical coherency) as a function of general cognitive-linguistic development and the acquisition of specific language skills during childhood. Consequently, children's PLAT scores should increase over time. Differences in the length of time children were allowed to write in the 1974 and 1975 versions of the PLAT make longitudinal comparisons of second versus third grade performance unfeasible for the available longitudinal subsample. However, a cross-sectional comparison of second with third grade performance in spring 1975 clearly indicates that children's written language does improve as expected from second to third grade on all variables (Table 4).

Correlations between PLAT scores and standardized aptitude and achievement test scores reported in Tables 4 and 5 pertain to both the criterion-related and construct validity of the PLAT. Correlations are reported separately for non-Follow Through and Follow Through, respectively. It was expected that:

- Moderate correlations would be found between PLAT performance and both aptitude and academic achievement as defined by existing standardized tests. Clearly the PLAT battery does not measure things entirely unrelated to what is measured by these other tests; in some degree it taps both cognitive and linguistic processes, including language mechanics, which underly performance on these other tests. On the other hand, PLAT scores were not expected to be entirely redundant with scores on these other measures since these measures do not assess the qualitative dimensions of divergent written language production.

- Stronger correlations were expected for the non-Follow Through than for the Follow Through sample. In the Cognitively Oriented Curriculum each child ideally has an opportunity to develop his/her

Table 4

PLAT Variables
Comparison of Second with Third Grade[1]
TOTAL SAMPLE
(N=900)

PLAT Variables	SECOND GRADE			THIRD GRADE			F ratio	Direction of Significant Effects
	Mean	*SD*	*N*	Mean	*SD*	*N*		
Fluency	40.44	*25.51*	*455*	58.42	*36.63*	*445*	90.04***	3rd > 2nd
Syntactic Maturity	7.74	*2.78*	*455*	8.32	*2.29*	*445*	12.70***	3rd > 2nd
Vocabulary Diversity	2.63	*.547*	*455*	2.98	*.644*	*445*	95.30***	3rd > 2nd
Descriptive Quantity	5.34	*5.24*	*455*	8.74	*7.47*	*445*	73.54***	3rd > 2nd
Descriptive Density	.109	*.066*	*455*	.126	*.062*	*445*	14.73***	3rd > 2nd
Descriptive Diversity	1.03	*.512*	*455*	1.33	*.593*	*445*	72.49***	3rd > 2nd
Descriptive Scope	2.25	*1.41*	*455*	3.07	*1.69*	*445*	73.02***	3rd > 2nd
Reporting Quality	2.39	*.872*	*439*	2.74	*.900*	*434*	35.66***	3rd > 2nd
Narrative Organization	1.92	*.558*	*441*	2.13	*.554*	*436*	31.60***	3rd > 2nd
Explanatory Statements	.208	*.528*	*455*	.435	*.725*	*445*	31.67***	3rd > 2nd
Decodability	.870	*.182*	*455*	.921	*.140*	*445*	22.78***	3rd > 2nd

*p < .05
**p < .01
***p < .001

[1]The grade level main effect was tested in a multiple linear regression design, covarying on treatment group, sex, and site.

Table 5

Correlation of PLAT Scores with Aptitude and Achievement Test Scores at Third Grade
Non-Follow Through

	MISSISSIPPI MISSOURI	FLORIDA		MISSISSIPPI			FLORIDA			NEW YORK CITY		MISSOURI	
	Binet	SFTAA		CAT			CTBS			NYCAT		ITBS	
Variable	IQ N=95-100	Verbal N=41	Nonverbal N=42	Reading N=39-43	Language N=38-41	Math N=38-41	Reading N=67	Language N=66	Math N=67	Vocab	Reading	Vocab N=49-51	Reading N=49-50
Fluency	.47	.32	.26	.55	.25	.48	.28	.27	.24	---	---	.43	.35
Syntactic Maturity	.18	.02	.02	-.08	-.05	-.17	.16	.17	.05	---	---	.18	.13
Vocabulary Diversity	.56	.32	.27	.50	.35	.32	.35	.37	.29	---	---	.43	.33
Descriptive Quantity	.60	.47	.28	.64	.41	.54	.41	.37	.33	---	---	.49	.40
Descriptive Density	.48	.50	.17	.27	.23	.24	.37	.34	.33	---	---	.35	.20
Descriptive Diversity	.66	.46	.15	.56	.44	.40	.39	.32	.34	---	---	.47	.37
Descriptive Scope	.64	.47	.29	.55	.43	.34	.42	.43	.41	---	---	.41	.28
Reporting Quality	.47	.40	.40	.35	.13	.12	.34	.29	.30	---	---	.37	.21
Narrative Organization	.38	.06	.16	.56	.40	.30	.18	.09	.29	---	---	.26	.26
Explanatory Statements	.45	.10	.24	.57	.20	.50	.07	.05	.17	---	---	.39	.26
Decodability	.31	.18	.20	.38	.21	.29	.35	.21	.28	---	---	.32	.35

particular strengths and is directly encouraged to do so. Consequently it is expected that different children will excel in different domains. Therefore, children in the High/Scope program should perform differently in different assessment situations, and the intercorrelations of outcome measures for children in this program should be generally lower than for children in more traditional classrooms. The presumption is made that children in traditional classrooms do not have the same opportunities to develop their individual strengths in ways of their own choosing, but rather that children who are conventionally bright and well socialized in the mainstream will generally do well in many domains while other children will perform at generally lower levels. As a result one would expect generally higher intercorrelations among diverse outcome measures.

Correlational findings tend to confirm these expectations. Tables 5 and 6 present correlations of PLAT variables with aptitude and achievement variables for non-Follow Through and Follow Through, respectively. Within each treatment group, the results are organized by site since different aptitude and achievement tests were used at different sites. With the exception of Syntactic Maturity, PLAT variables correlate moderately to strongly with Binet I.Q. scores in the non-Follow Through group while they correlate only weakly with Binet in the Follow Through group. Correlations between PLAT variables and Short Form Test of Academic Aptitude (SFTAA) scores tend to be moderate in both groups. This is perhaps explained by the fact that the SFTAA is administered in conjunction with the Comprehensive Tests of Basic Skills (CTBS) achievement test and is generally more highly correlated with achievement than is the Binet. Correlations between PLAT scores and California Achievement Test (CAT) subscores are substantially higher overall in the non-Follow Through than the Follow Through group. Correlations with CTBS achievement test scores tend to be somewhat higher in the non-Follow Through sample, but the difference is not as striking. Correlations with the New York City (Stanford) Achievement Test (NYCAT) are only available on Follow Through third graders; correlations are somewhat stronger within the Follow Through group than with the CAT and CTBS. Finally, PLAT scores tend to correlate moderately to strongly with Iowa Tests of Basic Skills (ITBS) scores within the Follow Through sample while relationships are generally less strong within the non-Follow Through sample. Overall, there is substantial evidence that PLAT performance is neither unrelated to nor redundant with aptitude and achievement particularly when the non-Follow Through sample is considered.

Table 6

Correlation of PLAT Scores with Aptitude and Achievement Test Scores at Third Grade Follow Through

	MISSISSIPPI MISSOURI	FLORIDA		MISSISSIPPI			FLORIDA			NEW YORK CITY		MISSOURI	
	Binet	SFTAA		CAT			CTBS			NYCAT		ITBS	
Variable	IQ N=164-165	Verbal N=40-41	Nonverbal N=40 41	Reading N=91-92	Language N=81-82	Math N=89-90	Reading N=77-79	Language N=77-79	Math N=77-79	Vocab N=30-32	Reading N=30-32	Vocab N= 49	Reading N= 49
Fluency	.08	.46	.47	.01	-.20	-.04	.35	.44	.28	.30	.37	.41	.28
Syntactic Maturity	.20	-.13	-.05	.29	.22	.31	-.04	-.11	-.13	-.01	.30	.33	.23
Vocabulary Diversity	.12	.58	.48	.14	.16	.21	.39	.49	.37	.29	.62	.52	.43
Descriptive Quantity	.19	.33	.28	.14	-.13	-.01	.27	.35	.24	.42	.48	.54	.48
Descriptive Density	.19	-.04	-.01	.22	.08	.08	.09	.11	.11	.50	.36	.65	.66
Descriptive Diversity	.17	.36	.38	.15	.07	.16	.28	.37	.30	.55	.58	.67	.58
Descriptive Scope	.24	.44	.32	.14	-.06	.07	.29	.39	.30	.43	.56	.58	.49
Reporting Quality	.21	.46	.24	.07	.03	.11	.32	.30	.20	.30	.16	.46	.55
Narrative Organization	.30	.39	.21	.01	-.03	-.08	.17	.17	.16	.19	.47	.24	.36
Explanatory Statements	.29	.14	.18	-.04	-.08	-.02	.12	.18	.17	-.15	.12	.46	.44
Decodability	.27	.27	.32	.19	.05	.07	.26	.26	.27	.34	.52	.40	.26

Evidence in support of the expectation that PLAT performance would be more weakly related to aptitude and achievement in the Follow Through sample than in the non-Follow Through sample is less strong but generally supportive. The differential correlations between PLAT and Binet scores across treatment groups at two sites are particularly striking. There was no expectation that PLAT scores would correlate more strongly with verbal than with nonverbal aptitude and achievement scales since these scales scores are highly intercorrelated among children in the study sample.

Finally, intercorrelations of PLAT variables are reported for the total sample and for second and third grades in Tables 7 through 9, respectively. Syntactic Maturity is essentially uncorrelated with other variables and, assuming its reliability, seems to measure an aspect of language production that is independent of other aspects. Decodability, Explanatory Statements, and Narrative Organization also appear to be relatively independent of other PLAT variables. Numerous strong correlations among other PLAT variables suggest that more global measures of fluency and quality might profitably be constructed. Decisions regarding the contruction of new variables await further developmental research.

Effects of Task Sequence

The 11 PLAT scores are derived from analyses of language samples produced in both the Reporting and Narrating tasks which, although similar in certain respects, differ with respect to stimulus materials and instructions to the child. In order to determine whether one sequence of task administration elicits higher levels of performance than the other, the relationship between task sequence and each of the PLAT subscores was analyzed. In almost all instances the Reporting task was administered before the Narrating task except in Mississippi where the sequence was systematically varied. In Mississippi, Follow Through and non-Follow Through children in second and third grades were randomly assigned to either the Reporting-Narrating or the Narrating-Reporting sequence. The effect of sequence on PLAT performance was tested in a multiple linear regression design, covarying on grade, sex, and treatment group. For the 283 children in the Mississippi sample, the sequence in which tasks were administered made no statistically significant difference in children's performance on any of the subscores of the PLAT.

Table 7

Intercorrelation of PLAT Variables
TOTAL SAMPLE
(N=900)

PLAT Variables	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
1. Fluency	1.00	.15	.74	.89	.36	.74	.79	.50	.36	.59	.44
2. Syntactic Maturity	.15	1.00	.30	.18	.16	.18	.18	.16	.17	.07	.22
3. Vocabulary Diversity	.74	.30	1.00	.69	.37	.79	.76	.46	.38	.46	.48
4. Descriptive Quantity	.89	.18	.69	1.00	.63	.81	.86	.55	.37	.57	.36
5. Descriptive Density	.36	.16	.37	.63	1.00	.71	.65	.50	.23	.22	.22
6. Descriptive Diversity	.74	.18	.79	.81	.71	1.00	.90	.60	.35	.45	.43
7. Descriptive Scope	.79	.18	.76	.86	.65	.90	1.00	.61	.39	.48	.45
8. Reporting Quality	.50	.16	.46	.55	.50	.60	.61	1.00	.28	.31	.39
9. Narrative Organization	.36	.17	.38	.37	.23	.35	.39	.28	1.00	.39	.25
10. Explanatory Statements	.59	.07	.46	.57	.22	.45	.48	.31	.39	1.00	.18
11. Decodability	.44	.22	.48	.36	.22	.43	.45	.39	.25	.18	1.00

Table 8

Intercorrelation of PLAT Variables
Second Grade
(N=455)

PLAT Variables	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
1. Fluency	1.00	.16	.67	.87	.33	.68	.75	.48	.27	.49	.48
2. Syntactic Maturity	.16	1.00	.34	.19	.16	.16	.18	.19	.17	.03	.22
3. Vocabulary Diversity	.67	.34	1.00	.60	.32	.74	.70	.39	.31	.36	.48
4. Descriptive Quantity	.87	.19	.60	1.00	.61	.75	.83	.50	.28	.44	.35
5. Descriptive Density	.33	.16	.32	.61	1.00	.69	.64	.43	.18	.12	.16
6. Descriptive Diversity	.68	.16	.74	.75	.69	1.00	.87	.53	.25	.30	.39
7. Descriptive Scope	.75	.18	.70	.83	.64	.87	1.00	.52	.30	.34	.45
8. Reporting Quality	.48	.19	.39	.50	.43	.53	.52	1.00	.17	.20	.38
9. Narrative Organization	.27	.17	.31	.28	.18	.25	.30	.17	1.00	.28	.23
10. Explanatory Statements	.49	.03	.36	.44	.12	.30	.34	.20	.30	1.00	.15
11. Decodability	.48	.22	.48	.35	.16	.39	.45	.38	.23	.15	1.00

Table 9

Intercorrelation of PLAT Variables

Third Grade

(N=445)

PLAT Variables	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
1. Fluency	1.00	.09	.76	.89	.37	.75	.79	.48	.37	.61	.39
2. Syntactic Maturity	.09	1.00	.23	.14	.15	.17	.15	.08	.13	.08	.18
3. Vocabulary Diversity	.76	.23	1.00	.71	.38	.81	.78	.47	.37	.48	.45
4. Descriptive Quantity	.89	.14	.71	1.00	.65	.83	.86	.57	.40	.61	.34
5. Descriptive Density	.37	.15	.38	.65	1.00	.72	.66	.55	.25	.27	.28
6. Descriptive Diversity	.75	.17	.81	.83	.72	1.00	.90	.63	.38	.50	.44
7. Descriptive Scope	.79	.15	.78	.86	.66	.90	1.00	.64	.42	.53	.43
8. Reporting Quality	.48	.08	.47	.57	.55	.63	.64	1.00	.33	.36	.37
9. Narrative Organization	.37	.13	.37	.40	.25	.38	.42	.33	1.00	.44	.25
10. Explanatory Statements	.61	.08	.48	.61	.27	.50	.53	.36	.44	1.00	.18
11. Decodability	.39	.18	.45	.34	.28	.44	.43	.37	.25	.18	1.00

It is tentatively concluded that children perform equally well on the PLAT regardless of the order in which they experience the two tasks. Occasional deviations from the Reporting-Narrating sequence at sites other than Mississippi will be disregarded in the analyses reported here in lieu of any indications that task sequence influences overall performance.

Effects of Interval Between Tasks

PLAT administrators scheduled testing to fit with building schedules, teacher schedules, and the availability of individual children. Considerable effort was made to administer both tasks to every child in the sample. As a result, there was wide, nonrandom variation in the number of days that elapsed between administration of the two tasks for different children (0 through 32 days). Reasons for variation in interval length are not entirely known but include: child absence, tester illness, difficult classroom schedules, and problems in coordinating simultaneous scheduling of tests in several schools distributed over a large geographic area.

The relationship of interval between tasks and PLAT performance was examined in order to address two questions: Does interval systematically influence performance? Is there an optimal interval for obtaining children's best performance?

Data from three sites were analyzed in order to address these questions: Mississippi, Missouri, and Chicago. Data from New York City and Florida were not included because interval distributions were badly skewed or because interval was systematically related to some independent variable(s). Since the intervals between tasks for most children fell between 1 and 7 days, it was decided to disregard those few children for whom length of interval was 0 days or over 7 days. The total sample was thereby reduced to 562 children (529 for PLAT variables Reporting Quality and Narrative Organization). The sample included 274 children from Mississippi, 131 from Missouri, and 157 from Chicago. The mean length of interval for the sample was 3.3 days, with a standard deviation of 1.9.

The interval effect was tested in a multiple linear regression design, covarying on grade, sex, treatment group, site, interval x site, and interval x group. The effects of interval on PLAT scores are summarized in Table 10. The length of interval between tasks had statistically significant

Table 10

The Effects of Interval (1-7 days) between Tasks on Children's PLAT Performance

PLAT Variables	Sample Size N	Proportion of Variance in PLAT Score Explained by Full Model[1] R_f^2	Effect of Interval: Magnitude of Effect in Variance $R_f^2 - R_r^2$	Effect of Interval: Direction of Effect[2]	Effect of Interval: Magnitude of Effect: Mean Increase in Score Per Additional Day of Interval (words)
Fluency	562	.27	.006*	positive	+2.63
Syntactic Maturity	562	.08	.000	---	
Vocabulary Diversity	562	.28	.003	---	
Descriptive Quantity	562	.24	.010**	positive	+ .72
Descriptive Density	562	.10	.009*	positive	+ .007
Descriptive Diversity	562	.25	.019**	positive	+ .09
Descriptive Scope	562	.26	.013**	positive	+ .21
Reporting Quality	529	.12	.002	---	
Narrative Organization	529	.10	.004	---	
Explanatory Statements	562	.19	.003	---	
Decodability	562	.09	.000	---	

*p < .05
**p < .01

[1]The full model contained the following independent variables: Interval, Grade, Sex, Group, Site, the interaction of interval with group, and the interaction of interval with site.

[2]The direction of effect is reported only when the effect of Interval reached significance at the .05 level.

and positive effect on 5 of 11 PLAT scores: Fluency, Descriptive Quantity, Descriptive Density, Descriptive Diversity, and Descriptive Scope. Two indices of the magnitude of the interval effect are reported. The average amount of gain in score per day increase in interval is perhaps more understandable than the proportion of variance accounted for; however, it is potentially misleading since the actual changes in mean scores by one-day intervals differ substantially from the values estimated by linear regression. The proportions of variance in PLAT scores predicted by interval are actually quite small (.006 - .019), indicating that relatively little of the variability in PLAT performance is explained by differences in the length of interval between tasks. Although plots of actual PLAT variable means by one-day intervals suggested that an interval of 5 to 7 days may be optimal for eliciting children's best performance, the advantage of one interval over another is slight.

In addition to these analyses, the average length of interval between tasks for Follow Through and non-Follow Through children in the sample was compared in a one-way analysis of variance. Follow Through children had significantly shorter intervals between tasks (mean = 2.9 days) than non-Follow Through children (mean = 3.9 days). Since length of interval was positively, albeit weakly, related to PLAT performance, the somewhat shorter mean interval in the Follow Through sample would seem to favor non-Follow Through in treatment group comparisons of PLAT performance.

FINDINGS: FOLLOW THROUGH/NON-FOLLOW THROUGH

Question: *Do Follow Through children write better reports and narratives than non-Follow Through children drawn from the same population?*

Expectations:

1. *At second grade, Follow Through children will not write better stories than comparable non-Follow Through children. Rationale: children are just beginning to write in earnest during second grade, and the impact of the Follow Through experience will not be detectable until children have had more opportunity to master basic writing mechanics.*

2. *At third grade, Follow Through children will write significantly*

better stories than comparable non-Follow Through children. Rationale: the Follow Through experience has a cummulative effect which becomes detectable by third grade.

3. *Differences favoring Follow Through are less likely to be reflected in the syntax than in the content of children's reports and narratives. Rationale: the Cognitive Curriculum encourages children to elaborate their thinking and to articulate increasingly complex thoughts through written language, rather than instructing children to write with more complex syntax. Although this educational process fosters the development of more complex syntax, it does so indirectly. Children acquire more complex syntax as it is needed for the expression of more complex thinking, rather than according to the grade level expectations embodied in most instructional materials.*

Findings: *Stated expectations were generally confirmed. By third grade, Follow Through children did indeed write significantly more complex and fluent reports and narratives than non-Follow Through children.*

Implications: *Follow Through children's superior ability to articulate their thoughts through written language bespeaks the program's effectiveness in achieving universal educational goals while at the same time satisfying curriculum requirements.*

Discussion

Findings from comparisons of Follow Through with non-Follow Through in the total sample are reported for second and third grades in Tables 11 and 12. The treatment group mean effect was tested within grade levels using a multiple linear regression analysis of variance design in which treatment group membership, sex, and site were used to predict PLAT

Table 11

PLAT Variables
Comparison of Follow Through with Non-Follow Through[1]
Second Grade

PLAT Variables	FOLLOW THROUGH			NON-FOLLOW THROUGH			F ratio	Direction of Significant Effects
	Mean	*SD*	*N*	Mean	*SD*	*N*		
Fluency	42.65	*25.99*	*277*	37.01	*24.43*	*178*	10.84**	FT > NFT
Syntactic Maturity	7.58	*2.60*	*277*	8.00	*3.02*	*178*	2.77	NS
Vocabulary Diversity	2.68	*.548*	*277*	2.54	*.535*	*178*	11.25***	FT > NFT
Descriptive Quantity	5.46	*5.53*	*277*	5.15	*4.75*	*178*	1.96	NS
Descriptive Density	.103	*.058*	*277*	.118	*.075*	*178*	5.03*	NFT > FT
Descriptive Diversity	1.06	*.513*	*277*	.985	*.510*	*178*	3.26	NS
Descriptive Scope	2.29	*1.42*	*277*	2.20	*1.40*	*178*	1.40	NS
Reporting Quality	2.29	*.760*	*276*	2.57	*1.01*	*163*	7.03**	NFT > FT
Narrative Organization	1.89	*.526*	*268*	1.96	*.604*	*173*	.649	NS
Explanatory Statements	.211	*.535*	*277*	.202	*.519*	*178*	1.11	NS
Decodability	.890	*.155*	*277*	.840	*.214*	*178*	7.18**	FT > NFT

*p < .05
**p < .01
***p < .001

[1]The treatment group main effect was tested within grade level in a multiple linear regression design, covarying on sex and site.

Table 12

PLAT Variables
Comparison of Follow Through with Non-Follow Through[1]
Third Grade

PLAT Variables	FOLLOW THROUGH			NON-FOLLOW THROUGH			F ratio	Direction of Significant Effects
	Mean	*SD*	*N*	Mean	*SD*	*N*		
Fluency	64.92	*39.07*	*270*	48.40	*29.97*	*175*	30.90***	FT > NFT
Syntactic Maturity	8.38	*2.20*	*270*	8.22	*2.43*	*175*	.735	NS
Vocabulary Diversity	3.12	*.625*	*270*	2.76	*.613*	*175*	41.44***	FT > NFT
Descriptive Quantity	9.39	*7.93*	*270*	7.74	*6.62*	*175*	10.22**	FT > NFT
Descriptive Density	.119	*.055*	*270*	.136	*.071*	*175*	5.82*	NFT > FT
Descriptive Diversity	1.38	*.589*	*270*	1.25	*.594*	*175*	6.84**	FT > NFT
Descriptive Scope	3.21	*1.69*	*270*	2.84	*1.66*	*175*	8.77**	FT > NFT
Reporting Quality	2.77	*.876*	*265*	2.69	*.939*	*169*	2.03	NS
Narrative Organization	2.17	*.535*	*267*	2.07	*.579*	*169*	4.91*	FT > NFT
Explanatory Statements	.419	*.723*	*270*	.460	*.728*	*175*	.043	NS
Decodability	.934	*.119*	*270*	.900	*.166*	*175*	6.99**	FT > NFT

*p < .05
**p < .01
***p < .001

[1]The treatment group main effect was tested within grade level in a multiple linear regression design, covarying on sex and site.

scores. The treatment group main effect was tested over and above the effects of sex and site. Interactions were not incorporated in the design because they were not of interest and because preliminary analyses indicated that they seldom made significant contribtuions to explanations of variance in the criterion measures.

Second grade findings (Table 11) are mixed. Follow Through children wrote longer stories (Fluency), used more diverse vocabularies (Vocabulary Diversity), and produced written text which was more readily decoded by adult readers (Decodability). However, non-Follow Through children obtained higher Descriptive Density and wrote reports which more nearly described how something was made (Reporting Quality). Overall, these findings tend to confirm the stated expectation of no difference between groups at second grade--i.e., neither group was clearly superior to the other when all findings are considered.

At third grade, however, PLAT findings clearly favor Follow Through over non-Follow Through (Table 12). Follow Through children write longer stories (Fluency), use more diverse vocabularies (Vocabulary Diversity), make more descriptive statements (Descriptive Quantity), have more diverse descriptive vocabularies (Descriptive Diversity), make more kinds of descriptive statements (Descriptive Scope), write better organized narratives (Narrative Organization), and produce written text which is more readily decoded (Decodability). On only one variable did non-Follow Through children score significantly higher than Follow Through children: Descriptive Density. No statistically significant differences between groups were found on Reporting Quality, Syntactic Maturity, and Explanatory Statements.

When these findings are considered together with the equally positive findings from 1973-74 research (Love and Bond, 1975), there seems little doubt that High/Scope's Cognitive Curriculum is effective in improving Follow Through children's ability to articulate their thoughts through written language. These findings have been obtained under a variety of field conditions in diverse geographical regions and among children representing different ethnic groups, in spite of apparent selection and instrument administration (interval) biases favoring the non-Follow Through comparison group. Consequently, it seems likely that the program effects obtained in High/Scope's Follow Through sample are generalizable to the larger population of children from low income families in the United States. Further support for the generalizability of the findings reported here and previously will be sought in High/Scope's continuing evaluation of the Cognitive Curriculum.

REFERENCES

Carroll, J. B. Language and thought. Englewood Cliffs, NJ: Prentice-Hall, 1964, p. 54.

Couvares, M. E., & Love, J. M. Assessing the writing of elementary school children: The development of a procedure and preliminary findings (Annual report to the Office of Education, Cognitively Oriented Curriculum Follow Through project). Ypsilanti, MI: High/Scope Educational Research Foundation, 1973.

Guilford, J. P. Psychometric methods. New York: McGraw-Hill, 1954.

Hunt, K. W. Grammatical structures written at three grade levels (Research Report No. 3) Champaign, IL: National Council of Teachers of English, 1965.

Love, J. M., & Bond, J. T. The High/Scope Productive Language Arts Test in the Cognitive Curriculum: Effects of Follow Through on the written language of second and third grade children (Addendum to annual report to the Office of Education, Cognitively Oriented Curriculum Follow Through project). Ypsilanti, MI: High/Scope Educational Research Foundation, March , 1975.

O'Donnell, R. C., Griffin, W. J., & Norris, R. C. Syntax of kindergarten and elementary school children: A transformational analysis (Research Report No. 8). Champaign, IL: National Council of English Teachers, 1967.

The "Examiner's Manual" and "Scoring Manual" have been removed at the request of the copyright holder. The Report with appendices is available through:

Research Department
High/Scope Educational Research
Foundation
600 North River Street
Ypsilanti, Michigan 48197

$3.00 each