ABSTRACT
          In the spring of 1975, the High/Scope Productive
Language Assessment Tasks (PLAT) were administered to virtually all
second-grade and third-grade children at five centers, who had been
enrolled in the Follow Through program since entering school; the
PLAT was also administered to groups of non-Follow-Through children
at four of five sites. Selection procedures matched children on grade
level, sex, ethnicity, economic status, and residential mobility;
nonetheless, at three of the four sites, the Follow-Through children
tended to come from more-disadvantaged homes than did the
non-Follow-Through children. Scores from the PLAT, which includes
both narrating and reporting tasks, indicated that, by the end of
third grade, Follow-Through children wrote more-fluent and
more-complex reports and narratives than did children enrolled in
non-Follow-Through classrooms. Detailed results are reported for each
of 11 PLAT variables. Analyses related to instrument validity are
also presented. (AA)

Bulletin of the High/Scope Foundation

Research Report: The Productive Language Assessment Tasks*

Design, methodology and findings from the 1974-75 assessment of second and third grade Follow Through children's written language production.

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 centers sponsored by the Foundation: Leflore County, Mississippi; Okaloosa County, Florida; P.S. 92, New York City; Howland and Lathrop Schools, Chicago; and Central Ozarks, Missouri. The PLAT was also administered to groups of non-Follow Through children at four of five sites; at one

*Formerly the Productive Language Arts Test

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. Careful examination of selection procedures suggests that, in spite of matching, Follow Through children tend to come from more disadvantaged homes than non-Follow Through children at three of the four sites having comparison groups. The PLAT was administered 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. 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 major purpose of collecting aptitude and achievement data was to further the development of the PLAT rather than to evaluate program effects.

THE PRODUCTIVE LANGUAGE ASSESSMENT TASKS are a measure of children's ability to express their thoughts and feelings through written language as this is influenced by their capacity to organize their experiences. It has two parts—a Reporting Task and a Narrating Task. In the Reporting Task, children are given a set of unstructured materials and are asked to make anything they want. Children may keep what they have made. After about 20 minutes, they are asked to write about "how they made" whatever they made. In the Narrating Task, children are given a set of relatively unstructured materials to help them make up a story. After about 15 minutes, they are asked to write a "make-believe or pretend" story. In both tasks, children are free to interact with one another during all parts of the tasks.

Existing standardized tests of language and reading achievement do not assess the productive language objectives of the High/Scope Cognitively Oriented Curriculum. Moreover, standardized language and reading achievement tests presume that all children should acquire specific language skills and specific vocabulary according to the same predetermined schedule, a presumption which contradicts the expectations for children in the Cognitively Oriented Curriculum at the primary grades.

2

3

The PLAT was designed to measure objectives that are central to the Cognitively Oriented Curriculum. The curriculum takes a nondidactic approach to language arts education that emphasizes generative language processes (oral and written expression) rather than language mechanics. Children use oral and written language to represent their own experiences and to express their own ideas and feelings. Reading builds upon writing experiences as children read what they have written or dictated themselves, then begin to read what others have written in order to obtain information or simply for pleasure. Teachers help children master standard English syntax, conventional spelling, punctuation, capitalization, and vocabulary as the need for these conventions emerges in the individual child's attempts to represent his own thinking in written language and to comprehend the written language of others. This approach to teaching the conventions of written language is highly individualized and frequently at variance with the grade-level norms of achievement tests.

PLAT variables

Initial processing of writing samples produced 32 major first-order variables, 16 for each task. Eleven second-order variables—those analyzed in this report and described below—were constructed from the first-order variable set during subsequent computer processing. All but two of these variables were derived by combining Reporting and Narrating scores. The purpose of combining the two writing samples was to obtain a more representative corpus of individual written language production and, thus, more reliable individual measures. Specific criteria for scoring are available from the High/Scope Foundation.

(1) *Fluency*—Fluency is an indicator of writing facility apart from any consideration of writing quality Computationally fluency = 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:

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

(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:

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

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, rationales, 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.

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.

For the purpose of calculating interscorer agreement, 15 samples of writing from the Reporting task and 15 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.

Scoring reliability was high; intraclass correlation coefficients exceeded 0.90 except for Number of Words in Extraneous Material when scored for Reporting stories where the coefficient was 0.89. 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 reliable. Though as yet there has been no formal instrument reliability study, there are indications that the PLAT will prove to be a reliable measure of an individual child's language as well. The evidence is adduced by reasoning backwards from findings pertinent to the assessment of instrument validity:

(1) When Follow Through children's performance at second grade was correlated with their performance at third grade, moderate stability in individual performance was found.

(2) Correlations between children's performance in the Reporting and Narrating tasks suggest instrument reliability. The Reporting and Narrating tasks are not alternate forms of the same test; rather, they are designed to tap presumably different dimensions of written language production. Consequently, it was not expected that performance on the two tasks 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. 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.

(3) Findings reported in the next section (Instrument Validity), indicating consistently higher performance by third graders compared with second graders and correlations of PLAT scores with aptitude and achievement test scores, suggest that the PLAT behaves like a reliable instrument.

Instrument validity

The PLAT has high content validity in that it samples the kind of language production encouraged by the Cognitively Oriented Curriculum and analyzes the language which children produce in curriculum-relevant ways. Moreover, responses from educators not associated with the High/Scope program suggest that the PLAT 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 1).

Correlations between PLAT scores and standardized aptitude and achievement test scores reported in tables 2 and 3 pertain to both the criterion-related and construct validity of the PLAT. It was expected that moderate correlations would be found between PLAT performance and both aptitude and achievement as defined by existing standardized tests. Clearly the PLAT does not measure things entirely unrelated to what is measured by aptitude and achievement tests; in some degree it taps both cognitive and linguistic processes, including language mechanics, which underlie performance on these tests. On the other hand, PLAT scores were not expected to be entirely redundant with scores on standardized aptitude and achievement tests since these measures do not assess the qualitative dimensions of written language production.

Correlational findings generally confirm this expectation. Tables 2 and 3 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 measures were used at different sites. With the exception of Syntactic Maturity, PLAT variables tend to have moderate to high correlations with aptitude and achievement measures in the non-Follow Through sample (table 2). Although correlation coefficients in the Follow Through sample (table 3) tend to be lower than in the non-Follow Through sample, it is clear that the PLAT is neither measuring something independent of nor redundant with aptitude and achievement tests.

In addition, it was expected that PLAT variables would tend to have higher correlation with aptitude and achievement measures in the non-Follow Through than in the Follow Through sample. The Cognitively Oriented Curriculum is a highly individualized approach to learning. Ideally each child has an opportunity to develop his particular strengths and is directly encouraged to do so. Consequently it is expected that different children will excel in different domains. This being true, children in the High/Scope program should perform differently in different assessment situations, and the intercorre-

Table 1

PLAT Variables:
Comparison of Second with Third Grade—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

*$p < .001$

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

4   8.

9

lations of outcome measures should be generally lower than for children in more traditional classrooms. The assumption is made that children in traditional classrooms do not have the same opportunities to develop their individual strengths but rather that children who are conventionally bright and well socialized in the mainstream will generally excel in most domains while other children will perform at lower levels regardless of their potential. As a result, one would expect higher intercorrelations among diverse outcome measures for children in more traditional classrooms.

The correlational findings presented in tables 2 and 3 offer some confirmation of this expectation. The differences in magnitude of correlations between PLAT variables and Binet scores across treatment groups in both Mississippi and Missouri are particularly striking. Correlations are substantially stronger in the non-Follow Through than in the Follow Through sample. Correlations between PLAT variables and CTBS scores in Mississippi also differentiate Follow Through from non-Follow Through in the expected direction. Correlations of PLAT variables with other measures of aptitude (SFTAA) and achievement (CTBS, NYCAT, ITBS) are inconclusive.

Table 2

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 IQ | SFTAA | | CAT | | | CTBS | | | NYCAT | | ITBS | |
Variable	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	No non-Follow Through group in New York		.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	.20	—	—	.26	.28
Explanatory Statements	.45	.10	.24	.57	.20	.50	.07	.05	.17	—	—	.39	.26
Decodability	.31	.18	.20	.38	.23	.29	.35	.21	.28	—	—	.32	.35

Findings: Follow Through vs. non-Follow Through

In general, Follow Through children were not expected to write better stories than non-Follow Through children at second grade. This expectation was based upon findings obtained in 1974 and upon the hypothesis that the Follow Through experience has a cumulative effect on children's writing ability which only begins to make itself felt at second grade and becomes much stronger by third grade.

The findings reported for second grade in table 4 are mixed. Follow Through children are more fluent, use a more diverse vocabulary and write more decodable stories than non-Follow Through children. However, non-Follow Through children have higher Descriptive Density and write reports which more nearly describe "how" something was made (Reporting Quality).

Third grade results were expected to more clearly favor children in the Follow Through sample. Findings presented in table 5 confirm this expectation. Although the syntax of Follow Through children is not significantly more mature than that of non-Follow Through children, Follow Through children are more fluent, use more diverse vocabularies, make more descriptive statements, 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).

Table 3
Correlation of PLAT Scores with Aptitude and Achievement Test Scores at Third Grade—Follow Through

Variable	MISSISSIPPI MISSOURI Binet IQ N=164-165	FLORIDA SFTAA Verbal N=40-41	Nonverbal N=40-41	MISSISSIPPI CAT Reading N=91-92	Language N=81-82	Math N=89-90	FLORIDA CTBS Reading N=77-79	Language N=77-79	Math N=77-79	NEW YORK CITY NYCAT Vocab N=30-32	Reading N=30-32	MISSOURI ITBS 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	.48	.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

Table 4

PLAT Variables:
Comparison of Follow Through with Non-Follow Through—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

Table 5

PLAT Variables:
Comparison of Follow Through with Non-Follow Through—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

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

On only one variable did non-Follow Through children score significantly higher than Follow Through children: Descriptive Density. No significant differences between groups were found on Reporting Quality and Explanatory Statements.

In sum, by the end of third grade, Follow Through children wrote more fluent and complex reports and narratives than children enrolled in non-Follow Through classrooms. This occurred in spite of indications that Follow Through children in the sample came from more economically disadvantaged homes than non-Follow Through children. The results reported here essentially replicate those obtained in 1973-74 research (Love and Bond, 1975). Together these findings strongly suggest that High/Scope's Cognitively Oriented Curriculum is more effective than traditional curricula in supporting the development of children's ability to communicate their thoughts through written language. The educational and social importance of this finding is clear. It reinforces High/Scope's commitment to an educational approach which involves each child as an active participant in the educational process rather than as a passive assimilator of conventional knowledge. Future research and development efforts will broaden High/Scope's language assessment activities to include oral as well as written language and to encompass an extended age range.   (J.T.B.)

References

Carroll, J.B. Language and thought. Englewood Cliffs. N.J.: Prentice-Hall. 1964.

Hunt, K.W. Grammatical structures written at three grade levels (Research Report No. 3). Champaign. Ill.: 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 (Report to U.S.O.E., Grant No. OEG-0-70-4919). 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. Ill.: National Council of Teachers of English. 1967.

Schools and research projects interested in using the High/Scope Productive Language Assessment Tasks as an assessment procedure, and desiring information on training to administer the tasks and the method of scoring, should contact:

John Love
Director, Research Department
High/Scope Educational Research Foundation
600 North River Street
Ypsilanti, Michigan 48197
(313) 485-2000

8

16