DOCUMENT RESUME

ED 145 922 PS 009 393

AUTHOR Becker, Wesley C.; Engelmann, Siegfried
TITLE Analysis of Achievement Data on Six Cohorts of Low-Income Children From 20 School Districts in the University of Oregon Direct Instruction Follow Through Model.
INSTITUTION Oregon Univ., Eugene. Coll. of Education.
SPONS AGENCY Bureau of School Systems (DHEW/OE), Washington, D.C. Div. of Follow-Through.
REPORT NO UO-TR-76-1
PUB DATE 76
GRANT G00-7507234; OEG-070-4257(286)
NOTE 239p.

EDRS PRICE MF-$0.83 HC-$12.71 Plus Postage.
DESCRIPTORS *Academic Achievement; Achievement Rating; Basic Skills; *Compensatory Education Programs; Demonstration Programs; *Disadvantaged Youth; Educationally Disadvantaged; Followup Studies; Low Income Groups; *Primary Education; *Program Effectiveness; Public Schools; Small Group Instruction; Teaching Techniques
IDENTIFIERS *Distar

ABSTRACT

This paper presents an analysis of data assessing the effectiveness of the Engelmann-Becker Direct Instruction Model (E-B) aimed at developing basic academic skills in disadvantaged primary-grade children (with the specific goal of enabling these children to achieve grade-level performance by the end of 3rd grade). The E-B model emphasizes small-group instruction, using sequenced daily lessons in reading, arithmetic and language designed by Engelmann and published under the trade name DISTAR. Data on almost 12,000 disadvantaged students in E-B programs in 20 Oregon school districts is presented in tables and accompanying text. Half the children had started the program in kindergarten, half in 1st grade. Achievement was measured by the Wide Range Achievement Test, the Metropolitan Achievement Test and the Slosson Intelligence Test. Results indicate model effectiveness in building basic skills for a wide variety of disadvantaged students. On all measures, children were at or near national norms by the end of 3rd grade. Follow-up testing of 5th- and 6th- graders demonstrated continuing effectiveness two and three years after the program in spite of losses against the norms. Testing losses in reading comprehension after 3rd grade are seen as due to the sudden shift to an unrestricted vocabulary in reading comprehension tests and 4th-grade textbooks. Report recommendations emphasize an early start to basic skill instruction, systematic expansion of basic vocabulary and the extension of follow-through programs until minimal adult competencies are met. (BF)

ANALYSIS OF ACHIEVEMENT DATA ON SIX COHORTS OF LOW-INCOME CHILDREN FROM 20 SCHOOL DISTRICTS IN THE UNIVERSITY OF OREGON DIRECT INSTRUCTION FOLLOW THROUGH MODEL

TECHNICAL REPORT 76-1

University of Oregon
College of Education
Follow Through Project
Eugene, Oregon, 97403

by

Wesley C. Becker and Siegfried Engelmann

This project was supported by funds from the U. S. Office of Education, Bureau of School Systems, Division of Follow Through Under Grant No. OEG-070-4257 (286) and GOO-7507234.

Table of Contents

The Direct Instruction Model*

Wesley C. Becker and Siegfried Engelmann
University of Oregon

The Direct Instruction Model emphasizes small-group face-to-face instruction by a teacher using carefully sequenced, daily lessons in Reading, Arithmetic, and Language. These programmed lessons were designed by Engelmann using modern learning principles and advanced programming strategies (Becker, Engelmann, and Thomas, 1975b) and are published by SRA under the trade name DISTAR. To achieve efficient teaching, the teacher needs only to concentrate on effective presentation techniques using the program materials.

In this report, we summarize the degree to which the Direct Instruction Follow Through Model has been effective by examining the data on nearly 12,000 disadvantaged students in E-B sponsored programs. For a fuller description of the model and the findings refer to Technical Report 76-1 and the references provided therein.*

Assumptions and Principles

The assumptions underlying the Engelmann-Becker Direct Instruction Model are:

1. All children can be taught.
2. The learning of basic skills is essential to intelligent behavior and should be the main focus of a compensatory education program.
3. The disadvantaged must be taught at a faster rate than typically occurs if they are to catch up with their middle-class peers.

We believe the goal of teaching all disadvantaged children basic skills at a faster rate can be accomplished by:

---

*This summary is adapted from: Becker, W. C. and Engelmann, S., Analysis of Achievement Data on Six Cohorts of Low Income Children from 20 School Districts in the University of Oregon Direct Instruction Follow Through Model. Technical Report No. 76-1, prepared for the Office of Education. Eugene, Oregon: University of Oregon, Follow Through Project, 1976. Program is funded by the U. S. Office of Education, Washington, D. C. 20202.

1. The use of instructional programs designed to teach the general case: When the teaching method can present only some examples in a set and the child learns to do all members of the set, a general case has been taught and an efficiency in teaching achieved. In the E-B Model, programs designed to do just that are provided to the teacher.

2. Increasing the manpower in the classroom. People are the primary tools of instruction. By adding teacher aides more instruction can occur.

3. Carefully structuring the daily program so that time is used to meet priority teaching needs. Everyone knows what to do and when to do it.

4. Using direct, small-group instruction. This is an efficient way to individualize instruction for the non-reader.

5. Using positive approaches to get and maintain student attention, reinforce good responses, and correct mistakes.

6. Using careful training and supervision to ensure that appropriate skills and procedures have been provided to the classroom staff.

7. Monitoring of student progress with biweekly criterion-referenced tests and reports of lesson taught to detect problems while there is time to correct them.

These component procedures have been used in designing a model program that will teach basic skills, if it is followed.

Program Objectives

Nine DISTAR programs are at the heart of the model. They serve to specify the teaching objectives for students in Reading, Arithmetic, and Language. In DISTAR Reading I and II, the focus is first on decoding skills and then comprehension. The children are taught to say the sounds for letters and to sound out regular words. Then irregular sound words and letter combinations are added. As story reading skills are built, the concern moves more to comprehension of what is read. In Reading III, the children are taught to read for new information and to use that

new information. Most of the stories have a science base to them which provides rules that can be used to solve problems in astronomy, muscle function, or measurement. Reading III should prepare the student to use upper level textbooks.

DISTAR Arithmetic I teaches basic addition and subtraction operations through a problem-solving approach. Number facts are memorized to speed up the process and to set the stage for more elaborate problems. In Arithmetic II, the students are introduced to multiplication and fractions. Addition and subtraction are extended to column operations, and a variety of measurement concepts involving time, money, length, and weight are taught. The students are also taught more number facts, how to derive unknown facts from known facts, and how to work story problems. In Arithmetic III, the students are taught algebra, factoring, and division, and the traditional operations are extended.

DISTAR Language I and II teach object names, object classes, object properties and relational terms. Children are taught to make whole statements and to appropriately and logically describe the world around them. Language I and II are basically comprehension programs, coupled with practice in language production. Among other things, the students learn to deal with conditionality, causality, multiple attributes, definitions, deductions, synonyms, and opposites. They are also taught how to ask questions in order to find out about something. Language III expands the logical use of language (reasoning) and teaches basic grammatical rules and structure. Many activities in Reading and Language are also geared to building writing and spelling skills.

The major goal of the Direct Instruction Model is to give disadvantaged children sufficient basic skills to compete with their more advantaged peers for higher education and the opportunities available in our society. As an immediate goal, we have used criterion of grade level performance (50th percentile) on major school achievement test batteries by the end of third grade. We are also

concerned that children learn arts, crafts, social skills and values, and we encourage instruction in these areas as suited to local conditions. We believe firmly that our procedures will build positive attitudes toward self, based on personal competence and positive treatment by others. However, a positive self-concept is viewed as a by-product of good teaching rather than as a goal that can be achieved in the abstract.

Installing the Program

The Contract

As a sponsor, we contract with a school district to establish and monitor a program they have chosen. The contract includes budgetary requirements for staff, materials, and facilities; job specifications for staff; training requirements; monitoring and evaluation procedures; and methods for dealing with contract violations. Initially, we trusted school districts to adhere to agreements. We did not insist on written contracts. Problems occurred when key personnel were changed without the new personnel being required to keep the agreements made with the sponsor. Now, we insist on acceptance of contract agreements before any person is hired to work in the program.

Training and Supervision

The goal of training is to provide the teachers (and parent aides) with the skills required to teach small groups within the model. Teachers learn how to group the children to get the most out of every child, how to use signals to coordinate group responding, how to present the DISTAR tasks, how to reinforce right responses, and how to correct mistakes. This training has usually been accomplished by a one- or two-week preservice workshop, continuing inservice sessions of about two hours a week, and classroom supervision. A number of detailed procedural manuals have been prepared for trainers and participants in training. With a structured teaching system, it is possible to specify the teaching skills

needed. Training involves modeling and actual teaching practice with teachers taking turns playing the students' roles. Teachers and aides learn by performing under supervision and gaining feedback on their performances, just as their students do. A videotape library which illustrates how to teach key tasks in the program has been helpful for inservice training. College credit is given for inservice courses in the basic reinforcement and programming principles underlying the program (Becker, Engelmann, & Thomas, 1975a, 1975b). This training has contributed to more than 50 of our aides becoming certified teachers.

Classroom supervision is provided by consultants trained by the sponsor. Many of these consultants are former teachers from the local schools. Every trainer must be able to demonstrate with children every aspect of the teaching requirements of the program. They are expected to spend 75% of their working day in the classroom.

Managers

Each project has a manager assigned by Oregon. Managers must be experienced teachers and supervisors who know the classroom procedures and assume responsibility for all phases of it. Managers adapt schedules to local needs, monitor all phases of the program, identify priorities, and continually work for improvement of implementation. Managers spend most of their on-site time in the classrooms.

Monitoring

The managers and supervisors provide the first line of quality control on teacher performance. The next line comes from biweekly reports of teaching activities and of student progress on tests. To monitor student progress, paraprofessionals are trained to give criterion-referenced tests in one area every two weeks. These biweekly reports show absences as well as where each group is in Reading, Arithmetic, and Language. The reports are used locally to regroup children, provide for tutoring, and/or to guide inservice training. They also let the sponsor know whether an acceptable rate of progress is being made, and the quality of that progress. Projections of yearly progress enable us to make program adjustments if the children

fall behind. Thus, the program has built into it feedback loops at all levels to provide a basis for corrective action at the time trouble occurs.

The Children Served by the Project

The main data described in this summary are based only on the Low Income students from 20 communities using the E-B Model. The communities include a cross-section of "poor" America--rural and inner-city Blacks, rural Whites, Mexican-Americans in Texas, Spanish-Americans in New Mexico, Native Americans in South Dakota and North Carolina, and a variety of ethnically mixed communities. Approximately 8,000 Low Income students are in the program at one time. The data to be presented cover students who entered the program between 1968 and 1973. We have data on 5,922 kindergarten-starting children and 5,565 first-grade starting children in our main analyses. Students are included in this analysis if they meet the OEO poverty guideline (called Low Income in this report), if they started the program at its earliest grade level, and if tests are available at more than one point in time. The data include students who may have left the program before completing it.*

For Wide Range Reading, the sample sizes are:

	Pre-K	Post-K	Post-1st	Post-2nd	Post-3rd
K-Starting:	2,435	5,181	4,810	3,300	1,988
		Pre-1st	Post-1st	Post-2nd	Post-3rd
1st-Starting:		2,412	5,160	4,665	3,629

*This data base provides a maximum sample size for measurement of program impact. To check for biases in the outcomes because of changes in students over grades (attrition), we have also analyzed year-to-year gains on the same students, and full-term pre-to-post gains (K-3 or 1-3) on the same students. These analyses do not materially change any conclusions except to make the actual gains about two-tenths of a grade level higher than those reported here. When the excluded non-Low Income children are added to the analysis (20% of our children), there is another slight increment in the level of performance. Children who enter the program late, perform a year lower on the average. This would be expected if the program is important.

The sample at the end of third grade is smaller because two groups that started in kindergarten and one group that started in first grade have not yet finished. The entry level testing is low because the first cohort was not pretested at all, and less than half of the next three cohorts were pretested. Examination of the sites with and without pretesting shows no detectable pattern that could introduce bias. Furthermore, a comparison of the partial data for cohorts 2, 3, and 4 with the latter full data on cohorts 5 and 6 shows no systematic differences in entry scores.

The Tests

Three tests were used. The Wide Range Achievement Test (WRAT) provides measures of Reading, Arithmetic, and Spelling. The Reading test is quite reliable (.91--.92) and has been demonstrated to reflect instruction which teaches decoding skills. The Arithmetic test has questionable content validity for some levels of the test and lower reliability (.72--.80), but provides a gross measure of computational skills. The spelling test has reasonable reliability and (.83) content validity. The Metropolitan Achievement Test (MAT) provides measures of Reading (comprehension and word knowledge), Math (computation, concepts, and problem solving), Spelling, Language (grammar), and at the Intermediate Level, Science. The MAT has excellent reliability (.93--.96 for the tests reported) and adequate validity as measures of some of our program goals. The Slosson Intelligence Test (SIT) is a short, individually administered test aimed at measuring what the Stanford-Binet Intelligence Test measures (reliability .92, correlation with Stanford-Binet .93). We have included it to measure some of the more general program goals tied to the logical use of language and reasoning skills.

The WRAT and the SIT have been given to nearly all students from the end of the first project year. The MAT was added at the end of first, second, and third grades beginning in Spring, 1972. Testing was carried out under the supervision of

the University of Oregon field staff by local teachers and aides. Careful training for and monitoring of testing was followed.

The Results

Norm-Referenced Gains

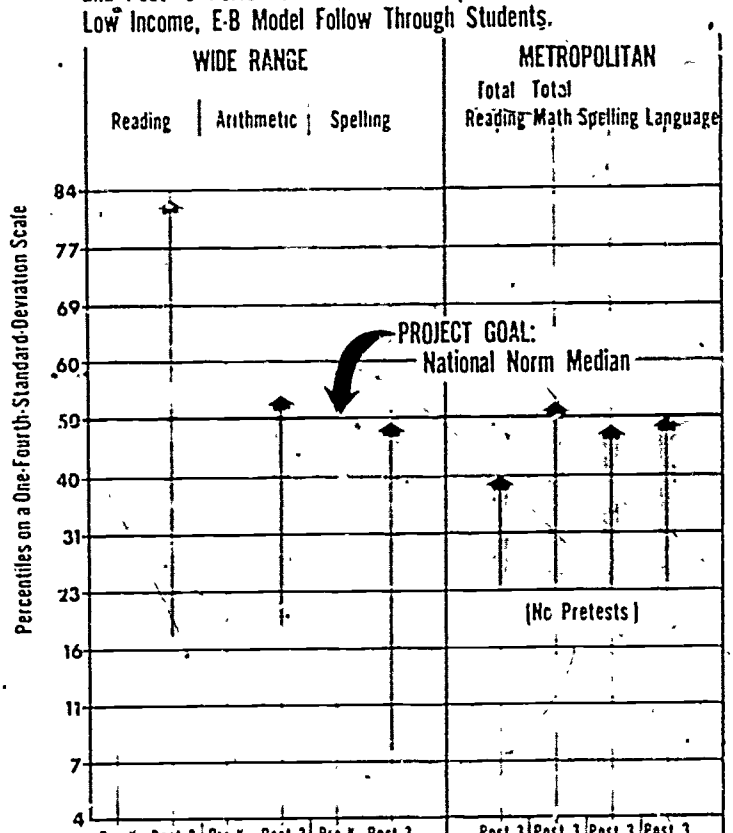
A major goal of the Direct Instruction Model has been to get Low-Income Follow Through students on a par with national norms by the end of third grade. Table 1 shows the degree to which this goal has been achieved for Low Income students starting the program in kindergarten.* On the WRAT substantial gains against the norms are present for all measures. These gains are displayed in Table 1 as percentiles on a standard-score scale.** The norm-referenced gain on WRAT Reading (decoding) is one-and-three-fourths standard deviation units. The Low Income students are close to a standard deviation above the norm on these important reading skills at the end of third grade. On WRAT Arithmetic and Spelling, large norm-referenced gains are demonstrated and the students are functioning at the national norm median.

We have no pretest for the Metropolitan Achievement Test measures. However, on the basis of the WRAT data we would estimate entry performance levels to be no higher than the 25th percentile and so we have used a jagged baseline to illustrate this probability in Table 1. The end of third grade performances show the Low Income E-B students to be at or near the national norm on all measures. They fall a little

*For the data in Table 1, statistics were computed with standard scores from tables provided by the test publishers. The mean standard scores were then converted to percentiles using the publishers' tables.

**Percentiles exaggerate differences near the mean and minimize those far from the mean. The graph has been drawn to overcome this problem. Assuming a normal curve, each horizontal line represents an increase of one-quarter standard deviation. However, the lines are labelled according to their percentile values. This provides the ease of interpretation of percentiles, as well as showing the magnitude of effects in standard score units. A one-fourth standard deviation gain is educationally important.

TABLE 1. Norm-Referenced Gains on the Wide Range (Pre-K to Post3) and Post-3 Performance on the Metropolitan for K-Starting, Low Income, E-B Model Follow Through Students.



	Pre K	Post 3	Pre K	Post 3	Pre K	Post 3	Post 3	Post 3	Post 3	Post 3
Number of Children	2435	1988	2676	1989	2520	1974	1863	1804	1824	1854
Percentiles	8	83	19	54	6	49	40	52	48	50
Mean Norm Group Based Standard Scores (SS)	85 9	114 5	86 6	101 2	79 5	99 7	58 0	71 1	62 8	69 9
Standard Deviation of SS	16 2	20 9	15 8	8 6	20 4	13 9	10 3	11 9	11 3	13 2
Mean Grade Equivalents	18	5 35	19	3 99	- 09	3 79	3.31	3 98	3 68	4 28

short of the norm (40th percentile) only on MAT Total Reading--a measure of reading comprehension.

Improvements over Cohorts

In K-starting sites the data for cohorts 2 to 6 show progressive improvements on WRAT Reading and Arithmetic (and to a lesser degree in Spelling). The magnitude

of the improvement is on the order of one-half standard deviation. The major improvements occur at the kindergarten level and are maintained in later grades. It took some kindergarten teachers time to take seriously the task of teaching academic skills to five-year-olds.

The data from 1st-starting sites show a narrowing of the gap with national norms, but with one less year to teach, the effects across tests are about one-quarter standard deviation below those shown for the K-sites in Table 1. There is a significant advantage in a compensatory catch-up program of having one more year to teach when that time is used efficiently.

Controlled Within-Project Comparisons

In 1970, we started the Cherokee, North Carolina program simultaneously in kindergarten, first grade, and second grade. This provided a basis for comparing the performance of the children who had progressively more of the model. Since 1970 we have also collected data on three additional cohorts, these data permit a further assessment of the progressive effects of better implementation (two of these cohorts have not yet finisned third grade).

The Cherokee project is located in the BIA school in the town of Cherokee. This is the only school on the reservation. With a constancy in student population and staffing, the study provides a rather clear-cut indication of program impact.

The results from this analysis at the end of third grade are given in Table 2. For each measure, Table 2 shows a progressive improvement with increased years of the program or longer implementation of the program. The magnitude of the improvements from early groups to later groups are in most cases more than one-half standard deviation. Only one of eight comparisons is not significant.

In six other sites, we started programs simultaneously in kindergarten and first grade. On nearly every comparison, those starting in kindergarten performed significantly higher on the achievement tests at the end of third grade. The

Table 2

End of Third Grade Percentiles for Cherokee Groups Having 2, 3, and 4 Years of the E-B Follow Through Model

	Two Years	Three Years	Four Years$_1$	Four Years$_2$
WRAT N	99	114	103	85
Reading	45th	79th	79th	84th
Arithmetic	34th	45th	55th	58th
MAT N	--	109	102	83
Total Reading	--	36th	47th	52nd
Total Math	--	47th	65th	68th
Language	--	40th	67th	67th
Spelling	--	36th	42nd	49th

Tests comparing mean standard scores for two- and four-year groups, and three- and four-year groups, were significant at the .05 level or better in each case except the three- and four-year comparison for WRAT Reading.

magnitude of the advantage averages, one-half standard deviation, or .6 to .8 grade levels. The overall findings are consistent with the previous comparison made between K-starting and 1st-starting sites, but the within-site comparisons show an effect twice as large. Because of regional variations in populations and school procedures between K-sites and 1st-sites, these within-site comparisons probably more truly reflect the potential gains of academic instruction during kindergarten for Low Income children.

Fifth and Sixth Grade Follow-Up

Approximately 700 E-B Follow Through fifth and sixth graders and 700 non-Follow Through comparison students were tested in seven sites at the end of fifth or sixth grade. The Wide Range Achievement Test (Levels 1 and 2) and the Metropolitan Achievement Test (Intermediate Level), were given in most cases. The comparison groups were from schools having students with similar backgrounds.

However, with samples of 35 to 100 students, variations in important background characteristics are to be expected. For this reason covariance analysis was used to adjust mean differences on outcome measures for differences in student sex, father's education, mother's education, number of siblings, income status, and ethnic group status. Four comparisons were made for most sites on the WRAT (Reading and Arithmetic using both levels of the test) and ten comparisons were made on the MAT (Word Knowledge, Reading, Total Reading, Language, Spelling, Math Computation, Math Concepts, Math Problem Solving, Total Math, and Science). The results showed 53 significant differences out of a possible 149 at the .05 level using a one-tailed test. Of the 53 significant differences, 50 favored Follow Through and 3 favored the non-Follow Through groups. Where significant effects are found, their magnitude averages about one-half standard deviation. WRAT Reading measures showed significant differences in favor of Follow Through in 14 of the 20 comparisons. MAT Reading measures showed 11 out of 39 significant comparisons in favor of Follow Through. WRAT Arithmetic showed significant comparisons favoring Follow Through in 4 of the 20 measures. MAT Math favored Follow Through significantly for 12 of 40 comparisons. Science showed 3 significant differences out of 10, Language 2 out of 10, Spelling 4 out of 10. One negative finding was found for WRAT Level 1 Arithmetic. The other two negative findings were on MAT Spelling. Spelling was not a strong point of the program in the early years.

Overall, the results are strongly supportive of an effect of the model which is persisting (in the absence of special programs) two and three years later. The results in Reading (both MAT and WRAT) are especially encouraging. The reader should note that these results were achieved largely with children from 1st-starting sites. Children from K-starting sites should do better.

The level of student performance in the years after Follow Through are in many cases disappointing. There were sizeable losses against national norms from third

to fifth, or third to sixth grade. Losses were especially great in math (about three-fourths of a standard deviation, 50th to 23rd percentile), and the reading comprehension (about one-fourth standard deviation, 35th to 25th percentile). Reading decoding skills were maintained, and apparently some schools taught spelling effectively in the intervening years. There is a clear implication that compensatory programs cannot be expected to maintain gains after the programs are stopped, unless all schools are prepared to provide effective teaching for every child in every grade. This is certainly not the case today.

Achievement of Students with Low IQ's

One of our concerns is the development of procedures for teaching the hard-to-teach student. A large group of such children can be defined by their low IQ's. We have examined the academic gains for our students with IQ scores below 80. Figure 1 shows comparisons for WRAT Reading gains for Low-IQ children and for All of our Follow Through children. The data show that Low-IQ children (mean IQ = 73 at pretest) gain more than a year on WRAT Reading for each year of instruction. On the average the gain is approximately 1.2 grade levels each year, while the gain for the All Group averages 1.35 grade levels each year. Gains in WRAT Arithmetic produce very similar results. The average gain for K-starting, Low-IQ students is .95 grade equivalents, while the All Group averages 1.00. For 1st-starting children, the average gain for the Low-IQ group is 1.04 grade equivalents and the All Group is 1.07.*

An examination of the number of lessons taught each year to the various groups shows that Low-IQ children are taught 30 fewer DISTAR lessons (160 is average) in kindergarten and 25 fewer in first grade. After that the differential drops to

*Examination of the gains in standard score units does not change the implications of the data.

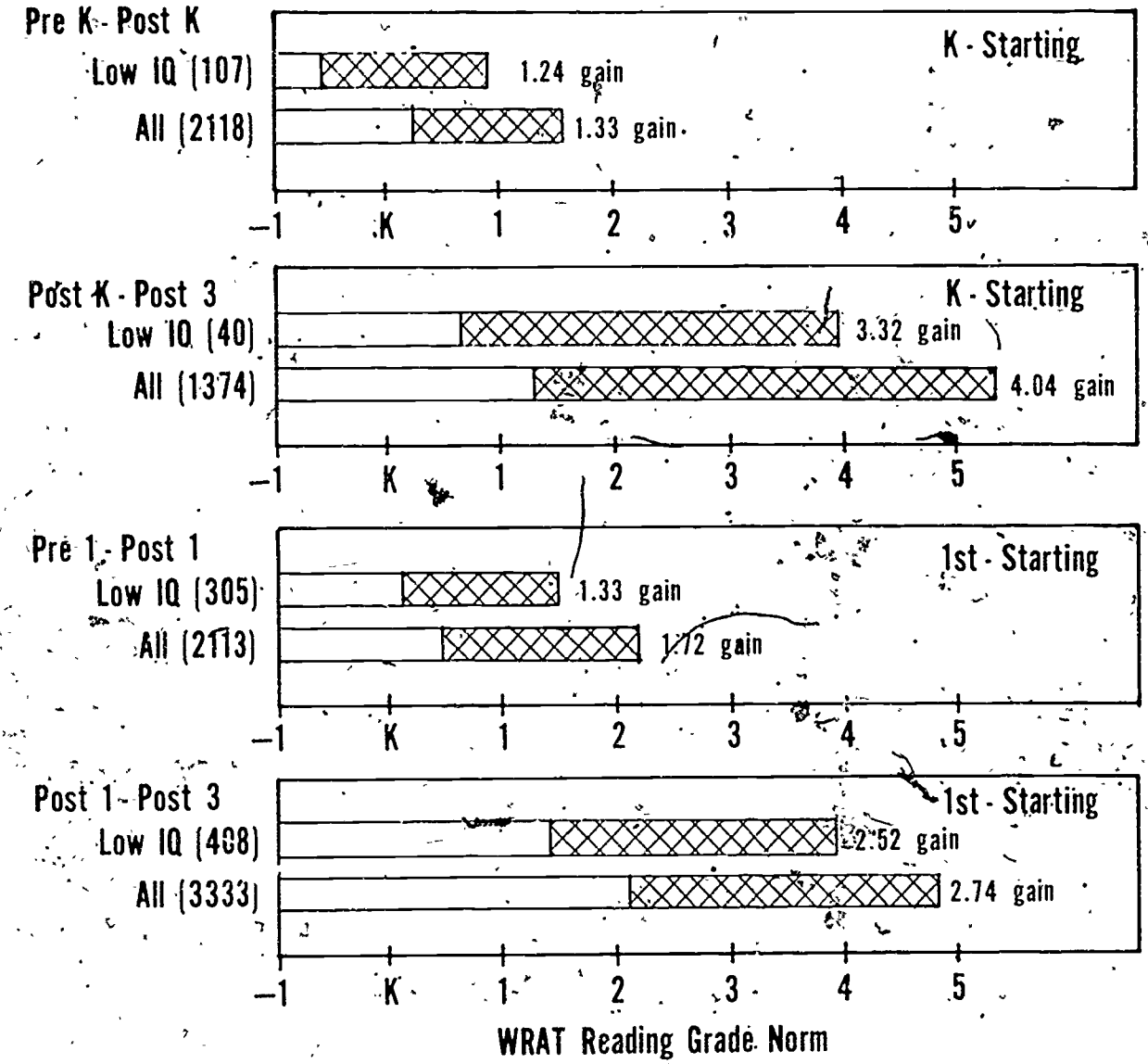


Figure 1 Reading gains of low IQ disadvantaged children. (Shaded area indicates gain for the time period shown to the left of each chart.)

Low IQ = IQ of 80 or less in Engelmann-Becker Follow Through Program.
All = All children in Engelmann-Becker Follow Through Program.

less than 15 lessons (or 9% fewer lessons). The data imply that Low-IQ children can be taught a lot more than is commonly assumed.

The children also showed gains in IQ. The All Group shows an average gain at the end of third grade of about eight points, both K-starting and 1st-starting groups. Discounting statistical regression effect,* the Low-IQ groups are estimated to have gained between 8 and 14 points on the average.

*The true score mean on pretest was estimated by using this formula $\bar{x}_T = \bar{x}_K \cdot r_{xx}$, where $\bar{x}_K$ is the mean of the Low-IQ group in deviation units, and r_{xx} is the reliability coefficient (.92).

Conclusions

1. The evidence shows that the model has been effective in building basic skills and intelligence for a wide variety of disadvantaged students. On all measures our Low Income students are at or near national norms by the end of third grade.

2. The evidence shows that the effects we have produced are educationally significant in size and can still be detected two and three years later. These significant follow-up findings have been shown largely with children having the benefit of only three years of the E-B program (1st-starting). K-starting students may do better. The significant fifth and sixth grade effects are found in spite of losses against the norms after leaving third grade.

3. The evidence strongly supports the value of beginning systematic instruction in basic skills early. There is a sizeable measurable benefit from Direct Instruction in kindergarten. Contrary to what has been attributed to our program by our critics, this has not been harmful to children. The National Evaluation measures of self-esteem (Coopersmith) and of parent and teacher attitudes (interviews) for those in our model show that our teaching approach produces very positive attitudes as well as academic gains. Our students are enthusiastic youngsters who enjoy learning and enjoy becoming competent.

4. The evidence demonstrates that the Direct Instruction Model is clearly effective with Low-IQ children. This is the group which has failed the most in the past.

5. The evidence shows systematic improvement for successive K-starting groups. Better implementation of the program, as well as program improvements, have led to better outcomes. Not yet seen in the data are the additional benefits likely to ensue from the second-generation of the DISTAR programs, which have been revised in the last two years.

Further support for these conclusions is provided by Rosenshine (1976). In his review of research on classroom instruction, Rosenshine concludes that "in the current research findings on time, content covered, work groupings, teacher questions, student responses, and adult feedback, one sees a general convergence of results on what might be labeled the direct instruction model (sometimes called a structured approach)." Rosenshine draws on the research of Stallings and Kaskowitz (1974), Soar (1973), and Brophy and Evertson (1974) in drawing his conclusions. Where the components of the direct instruction model described earlier in this report are present, positive gains in Reading and Math are found.

Discussion

Two points in the data are expecially in need of discussion. Why are we able to reach national norms on all tests except reading comprehension, and why do the losses against the norms occur from grades three to five or three to six? The answers to these questions are related.

It is our conviction, based on a variety of evidence, that a good proportion of children from Low Income homes do not receive the instruction in basic language concepts that the typical middle-class child receives. Since basic vocabulary is a key element to success on intelligence tests, these children are likely to test lower on such tests. To a large extent, schools fail to systematically build basic vocabularies in students. Vocabulary in readers is controlled for the first three grades (and includes only about 1500 words according to Chall) and then suddenly shifts to a unrestricted adult vocabulary of about 15,000 words (Thorndike-Lorge estimate). This sudden shift is reflected in the performance of our students on the Metropolitan Elementary Level Test at the end of third grade. Reading comprehension is tested with a vocabulary that goes far beyond what has been typically taught in school. This same vocabulary is also present in most fourth grade textbooks. It is not surprising that children of poverty are likely to flounder at this time, or

that our students should show losses between the end of third grade and sixth grade when there is no Follow Through. The Low Income students do not get the home instruction in Language that middle class children get.

The losses in Arithmetic between third and sixth grade simply imply that school programs are not adequate (do not teach) for the instruction of Low Income students at this level. The students in our programs showed they could learn when they were taught.

If this analysis is correct, the implications are:

1. Begin basic skill instruction no later than kindergarten for all educationally disadvantaged children.

2. Revise school programs to systematically build Language comprehension through a gradual expanding of the admissible basic vocabulary. Do not leave this basic competency up to the home or chance.

3. Extend a Follow Through type program through the school years until at least minimal adult competencies are met.

References

Becker, W. C., Engelmann, S., & Thomas, D. R. Teaching 1: Classroom Management. Palo Alto: Science Research Associates, 1975a.

Becker, W. C., Engelmann, S., & Thomas, D. R. Teaching 2: Cognitiv Learning and Instruction. Palo Alto: Science Research Associates, 1975b.

Brophy, J. E., & Evertson, C. M. Process-Product Correlations in the Texas Teacher Effectiveness Study: Final Report. Austin, Texas: The University of Texas, 1974.

Rosenshine, B. Classroom instruction. N. L. Gage (Ed.) The Psychology of Teaching Methods. NSSE Yearbook, 1976. Chicago: University of Chicago Press (in press).

Soar. Follow Through Classroom Process Measurement and Pupil Growth. Austin, Texas: The University of Texas, 1974.

Stallings and Kaskowitz. Follow Through Classroom Observation Evaluation. Austin, Texas: The University of Texas, 1974.

CHAPTER 1

THE MODEL, ITS GOALS, AND THE TESTS

Follow Through was initiated when it appeared that gains made in some effective Head Start programs were being washed out in the early school grades. Originally planned as an operational program, it was converted to a planned variation experiment when the initial funding level was cut from $120 million to $15 million. While authorized under the Economic Opportunity Act, the program was administered by the Office of Education from the start. A pilot program was started in 1967-68 in 40 school districts. During this year, the decision to find out what works was made, and "program sponsors" were sought. Each Follow Through community was asked to select from a set of predeveloped approaches the one they would like to work with.

The initial group of potential sponsors included Glen Nimnicht, David Weikart, Ira Gordon, Leonard Sealey, Marie Hughes, Don Bushell, Larry Gotkin, and Siegfried Engelmann. Consideration was given to the possibility of combining some sponsors into consortia. For example, the University of Oregon Direct Instruction Model, the University of Pittsburg Individualized Early Learning Model, and the University of Kansas Behavioral Analysis Model all relied on modern behaviorism for basic principles. However, the differences between approaches were considerable. Pittsburgh and Oregon were developing their own materials. One used individualized instruction and the other small-group instruction as the main vehicle. Kansas relied almost entirely on published materials. Thus, the idea of consortia was abandoned.

After preliminary discussions, 18 sponsoring groups were invited to present proposals to Follow Through communities for consideration. Formally sponsored programs were begun in the summer of 1969. Eventually, 165 school districts and 22 sponsors came to be involved.[1]

By 1970, we at Oregon were working with 20 school districts, starting at the beginning level and adding a grade a year until the students were in third grade. A new group of beginning level students has been added to the study each year since 1968. In a given year, more than 9,000 students are in the Oregon sponsored program.

The Follow Through Guidelines call for a program of comprehensive services including health care, social and psychological services, community involvement, and an educational program. Sponsors could be responsible for any or all of these services. In most cases, however, the communities assumed responsibility for health and nutrition programs. The focus in what follows is on the system built to install and operate the educational program.

BRIEF DESCRIPTION OF UNIVERSITY OF OREGON

Direct Instruction Model[2]

Engelmann defined the basic problem faced in teaching disadvantaged children as one of devising a system to get more teaching going in the classroom. Only if disadvantaged children were taught more could they learn more.

Our Follow Through instructional system was developed having these components:

1. Increased manpower in the classroom.
2. Structured daily routine.
3. Daily programmed lessons.
4. An efficient teaching method.
5. Continuing training.
6. Monitoring of progress of the children and the skills of the teachers.

1. Increased manpower in the classroom. When children cannot read, the primary means available for instructing them is by talking to them. If one is to get to every child and fully utilize the school day for instruction at a faster than average rate, more than one teacher is required for 25 to 30 children. Because of cost considerations, two teacher aides were used. For the most part the aides are parents of the poor children. It was our belief that parents who learned good teaching skills would also be in a better position to facilitate their children's learning at home.

2. Structuring the daily program. Manpower by itself does not insure that more teaching goes on in the classroom. The organization of the school day, a good program, and training are needed to effectively use the added manpower. The classrooms are set up so that the three "teachers" are each working in booths (for sound control) with groups of four to seven children. The teachers and aides become specialists in one of the three basic programs (Reading, Language, and Arithmetic) and a schedule is devised to fit each school's timetable to rotate the children through teaching groups and other activities when the children work on their own. Approximately thirty minutes is used for small group instruction in each subject area at Level I and II. At Level III, 15 minutes of instruction is followed by thirty minutes of self-directed practice in workbooks.

3. Programmed lessons. The instructional programs that are used in our Follow Through classrooms are the DISTAR® programs (Reading, Arithmetic, and Language). These programs are potentially powerful--particularly with respect to teaching the general case.

4. An efficient teaching method. The DISTAR programs are just words on paper. In order to teach these skills, the teachers and aides must understand the concepts and operations they are teaching and must have a number of basic teaching skills. These skills involve management of the children and organization of the teaching materials so that both the children and the teacher are ready to work when they sit down in an instructional group. Beyond that, the teacher needs to know how to teach a task--any task.

The skills involved consist of knowing how to present the demonstrations to the children; how to use attention signals to get the children to respond together (or individually) on cue; how to pace each task appropriately, quickly enough to hold attention, yet going slowly when required to give the children "time to think"; how to use reinforcers effectively to strengthen correct responding; and how to correct mistakes in a way which permits all children to learn each task (criterion teaching).

5. Training and supervision. The goal of training is to provide the teacher with the skills outlined above. This is accomplished in a one or two-week preservice workshop, continuing inservice sessions of about two hours a week, and through classroom supervision. A number of detailed procedural manuals have been prepared for trainers and participants in training. The key is to know what the teachers should be able to do, and to devise procedures to teach the required skills. It should be recognized that precision in specifying and training essential teaching skills is only possible within a structured teaching system.

Classroom supervision is provided by consultants trained by the sponsor. Many of these are former teachers from the local site. There is approximately one local supervisor for every two hundred children in the program.

6. Monitoring. The management of the progress of more than 9,000 children in 20 locations around the country requires a carefully designed monitoring system.

Built into the DISTAR programs are teacher-given tests to check each new skill as it is taught. To monitor child progress independently of the teacher, continuous progress tests (criterion referenced) are given in each area each six weeks by paraprofessionals at the Follow Through sites. Every two weeks test results in one area are summarized by child on four-copy IBM forms, (with names and numbers preprinted by group). These biweekly reports also show absences for the two-week period and show where each group is in each program. Copies of the reports go to the teacher, the supervisor, the Follow Through Director, and our data analysis center. The reports can be used locally to directly regroup the children or to provide special remediation or acceleration. They also provide a basis for summary analyses of progress for management by the sponsor. Trouble spots can be determined and worked on.

Management reports are produced by computer to keep track of group progress. Projections are made and compared with target goals for each group for the year. When projections fall behind goals, adjustments in the program can be made at the site to attempt to reach goals before it is too late to do anything about it. Management reports also keep track of school calendars and absences so that it is possible to base projections for each site on local conditions that affect teaching days available.

PARTICIPATING FOLLOW THROUGH PROJECTS

This report covers data collected in the school years 1968-69 to 1974-75 in 20 Follow Through Projects sponsored by the University of Oregon, Direct Instruction Model. The sites, entry levels and years covered by this report are shown in Table 1.1.

Table 1.1

Sites by Years and Entry Level*

Sites	School Years						
	68-69	69-70	70-71	71-72	72-73	73-74	74-75
Brooklyn, NY, PS137	K	K	K	K	K	K	K
Cherokee, NC			K-1	K	K	K	K
Chicago, IL (Ogden)		K	K	K	K	(not Title I eligible)	
Dayton, OH	1	1	1	1	1	1	1
Dimmitt, TX		1	1	1	1	1	Dropped out in 2/75
E. Las Vegas, NM	1	1	1	1	1	1	1
E. St. Louis, IL	1	1	K-1	K	K	K	K
Flint, MI		K	K	K	K	K	K
Flippin, AR		K-1	K	K	K	K	K
Grand Rapids, MI	K	K	K	K	No Agreement		Dropped
Providence, RI		K-1	K	K	K	K	K
Racine, WI	K-1	K	K	K	K	K	K
Rosebud Tribe, SD	1	K	K	K	K	K	K
Smithville, TN	1	1	1	1	1	1	1
Todd County, SD	1	K	K	K	K	K	K
Tupelo, MS	1	1	1	1	1	1	1
Uvalde, TX	1	1	1	1	1	1	1
Washington, DC	K	K	K	K	K	K	K
W. Iron County, MI	K	K	K	K	K	K	K
Williamsburg County		1	1	1	1	1	1

*K = K-Starting, 1 = 1st-Starting

OBJECTIVES OF THE EVALUATION

The evaluation was designed:

1. To demonstrate that economically disadvantaged children can be taught so that they will show a rate of progress typical for national norm groups in reading and math for grades K to 3. (Priority: high)

2. To demonstrate that economically disadvantaged children can be taught so that they show a mean IQ gain from entry grade through third grade, even though the usual expectancy is for a loss on IQ tests. (Priority: middle)

3. To demonstrate that low IQ children (80 and below) from economically disadvantaged homes can be taught at least a grade level a year. (Priority: high)

4. To determine if any found gains can be shown to persist at grades 5 and 6.

5. To examine the effects of introducing the model into Head Start.

6. To show the possible relations of process and outcome measures.

Designs

A norm-referenced design will be used as one approach to objectives 1, 2, and 3.

Control-groups designs are used for objectives 1, 2, 4, and 5.

Experimental approaches are used for objective 6.

Measures

1. Wide Range Achievement Test (WRAT)

   (a) Reading (decoding words) (R)
   (b) Arithmetic (A)
   (c) Spelling (S)

2. Metropolitan Achievement Test (MAT)

   (a) Primary I.

      (1) Word Knowledge (WK)
      (2) Word Analysis (WA)
      (3) Reading (R)
      (4) Total Reading (TR)
      (5) Total Math (TM)

   (b) Primary II.

      (1) Word Knowledge (WK)
      (2) Word Analysis (WA)
      (3) Reading (R)
      (4) Total Reading (TR)
      (5) Spelling (S)
      (6) Language (L)
      (7) Math Computation (MCom)
      (8) Math Concepts (MCon)
      (9) Math Problem Solving (MPS)
      (10) Total Math (TM)

   (c) Elementary (Same scores as Primary II less Word Analysis.)

   (d) Intermediate (Same as Elementary plus Science.)

3. Slosson Intelligence Test (SIT).

4. Attendance.

5. Lessons taught in programs (Reading, Arithmetic, and Language).

6. Criterion-referenced tests (Reading, Arithmetic, and Language) for the years 1970 to 1973.

7. Teacher-aide background form.

8. Student background form.

9. Individual site parent and teacher questionnaires.

10. Teacher and aide evaluations by supervision.

11. Student performance measures in experimental studies of instruction.

RELATION OF OBJECTIVES TO MEASURES

Objective	Measures	In National Evaluation	In Site Evaluations	Appropriateness of Measure (5-very appropriate 1-barely suited)
1. Norm-Referenced Gains:				
(a) Reading	WRAT-Reading	(was)	Yes	5 (decoding)
	MAT-WK, WA, R, TR	Yes	Yes	4-levels 1, 2 3-level 3 (Comprehension)
(b) Math	WRAT-Math	(was)	Yes	4-levels 1, 3 2-level 2
	MAT-MCom, MCon, MPS, TM	Yes	Yes	4-levels 1, 2, 3
(c) Supplemental Measures	WRAT-Spelling	No	No	3
	MAT-Spelling	Yes	Yes	3
	MAT-Language	Yes	Yes	5
2. IQ Gains:	SIT	No	Yes	4
3. Instruction of Low IQ Students:				
(a) Reading	WRAT-Reading	(was)	Yes	5
	MAT-Reading	Yes	Yes	3
(b) Math	WRAT-Math	(was)	Yes	4-levels 1, 3 2-level 2
	MAT-Math	Yes	Yes	4
4. Fifth and Sixth Grade Study:				
(a) Reading	WRAT-Reading			5 (decoding)
	MAT-Reading			4 (comprehension)
(b) Math	WRAT-Math			4
	MAT-Math			4
(c) Supplemental Measures	MAT-Science			4
	MAT-Language			3
	MAT-Spelling			3
5. Effects of Head Start (Planned Variation):				
(a) Reading	WRAT-Reading			5 (decoding)
	MAT-Reading			4 (comprehension)

Objective	Measures	In National Evaluation	In Site Evaluations	Appropriateness of Measure
(b) Math	WRAT-Math			4
(c) Supplemental Measures	MAT-Math			4
	SIT-IQ gain			4
	MAT-Spelling			3

6. Relations of Process and Outcome:

Experimental Studies	Student Rate and Accuracy	No	No	?
	Time on Task	No	No	?
	Amount Mastered	No	No	?

RELIABILITY OF STANDARDIZED MEASURES

Wide Range Achievement Test (WRAT)

From the test manual:

Age Group	N	Split-half Reliability Reading	Arithmetic	Level I Spelling
5	200	.981	.966	.971
6	200	.986	.959	.963
7	200	.993	.962	.977
8	200	.991	.948	.978
9	200	.989	.942	.977
10	200	.990	.948	.981

These are probably somewhat inflated because of the ten-second time limit per item or because of the time limits on test sections. However, the time limits are very generous and SRI alpha coefficients on a combination of reading and math items ranged from .93 to .97 for various samples.[3]

The ABT report on Follow Through data for 1969-1972 gives internal consistency coefficients for partial WRAT tests for a random sample of the Follow Through test population (N = 6401) as follows:[4]

	Pre-K Testing	Post-K Testing
Reading	.91	.92
Arithmetic	.63	.72
Spelling	.83	.83

How much these estimates are reduced by the shorter tests used is not known. However, the data are close to what we have found for similar groups on the full tests.

Our analysis of a sample of 50 records for students pretested in the fall of 1970 and posttested in the spring of 1971, and the spring of 1972 showed the following alpha coefficients:

	Reading	Arithmetic
K Pretest Fall, 1970	.92	.79
K Posttest Spring, 1971	.94	.80
1st Posttest Spring, 1972	.94	.76

The lower internal consistency reliabilities for WRAT Arithmetic as used in our program are partly attributable to its lack of sensitivity to what we teach. This reduces the range of scores, particularly at level II.

Slosson Intelligence Test (SIT)

The test manual states that test-retest reliability is estimated to be .92 for a population with standard deviation of 15.

Metropolitan Achievement Test (MAT)

The test manuals report the following alpha reliabilities:

Scales	Test Level			
	Primary 1	Primary II	Elementary	Intermediate
Word Knowledge	.88	.93	.94 (.94)*	.92
Word Analysis	.90	.90	--	--
Reading	.95	.93	.92 (.88)	.93
Total Reading	.96	.96	.96	.96
Spelling	---	.94	.96 (.97)	.90
Language	--	--	.93 (.89)	.95
Math Computation	--	.86	.88 (.89)	.84
Math Concepts	--	.85	.90 (.89)	.88
Math Problem Solving	--	.88	.91 (.88)	.89
Total Math	.93	.95	.96	.95
Science	--	--	--	.94

*From ABT Report, Vol. III[5]

VALIDITY INFORMATION ON STANDARDIZED MEASURES

Wide Range Achievement Test (WRAT)

Interpretability of Grade Norm Scores. Horst et al.[6] have indicated that grade equivalent scores for most norm-referenced tests are difficult to interpret because kids at different grade levels do not take the same test, and in relation to standard scores, the amount of learning represented by a grade-norm change of one year can vary drastically for different grade levels. For example, in terms of standard scores, a grade-equivalent change from 2.0 to 4.0 on the MAT Total Reading Test is the same as a change in grade-equivalents from 5.0 to 9.0.

The first criticism is not appropriate to WRAT Reading scores, since all students from age 5 to 12 take the same test. However, there is a progressive change in the relation of standard scores to grade-equivalents. In progressing from grade 1 to 7, the standard deviation in grade-equivalent units increases from 1.00 to 2.25 grades. This implies that the grade-equivalent score changes get relatively smaller in standard score units with a progression in grade units. Because of this problem, statistical analyses will usually be based on standard score or raw score units, and grade equivalents will be used primarily as an aide to the interpretation of the results.

Sensitivity to instruction. We have published several analysis showing systematic relations between progress on the WRAT Reading and Arithmetic tests and progress in the DISTAR Reading and Arithmetic programs for levels I to III.[7,8] The data imply that the WRAT Reading Test is fairly sensitive to instruction in the DISTAR programs (a valid measure of one goal of the program--decoding skills). The WRAT Arithmetic Test is also fairly sensitive to the program at levels I and III, but not to level II. The lack of sensitivity at level II is due to the DISTAR program focusing on problems presented in rows (important for later work in algebra), while the WRAT has a high percentage of problems presented in columns. At the end of level III, this is no longer a problem, but the test again becomes insensitive at the fifth and sixth grade levels. Within year-level gains on WRAT Reading are significantly correlated with number of lessons taught in Distar Reading I and II.

Slosson Intelligence Test (SIT)

The Slosson Intelligence Test is a short, easily scorable, test designed to measure what the Stanford-Binet measures. For groups at

each age from 5 to 12, the correlations with the Stanford-Binet Form LM were found to range from .94 to .98. However, the standard deviation for the groups tested were between 18 and 24, rather than 15. This would serve to inflate the correlations. With appropriate corrections, the average validity coefficient is about .93.

Metropolitan Achievement Test (MAT)

The basic statement about the validity of the MAT is contained in the following quote from the Teacher's Handbook:

> The validity of an achievement test is defined primarily in terms of content validity. A test has content validity if the test items adequately cover the curricular areas that the test is supposed to evaluate. Since each school has its own curriculum, the content validity of Metropolitan Achievement Tests must be evaluated by each school. It cannot be claimed that the tests are universally valid. To assist schools in judging the content validity of the tests, the authors and publisher have prepared content outlines for the tests and described the procedures used in developing the test content.

As this quote suggests, the validity of any achievement test must be judged by examining its content in relation to the program of instruction. We have examined the MAT at its various levels in relation to the DISTAR programs and find the test to measure many things that the programs do not teach, and not measuring much that the programs do teach. The major problem that arises, is that reading skills (decoding and comprehension) are tested (especially at the Elementary Level--end of third grade) using a vocabulary the students have not been taught. As one moves from the Primary I and Primary II Levels of the MAT to the Elementary Level, reading becomes more and more defined as comprehending a vocabulary of unrestricted range (e.g., a full adult vocabulary). The children have to understand (as well as decode) stories about an

English holiday (Guy Fawkes Day) and the "Gunpowder Plot" against the government, they have to read about Amazon Ants, Alexander Fleming's discovery of penicilin, architecture as an art form, museums with Egyptian artifacts, and a seated cat. Then they often have to answer questions based on conjecture about things not actually read about.

There are reasons to question the validity of MAT Elementary Level Reading as a measure of the instruction occurring in our program.

The following content descriptions of the MAT tests are taken from the MAT Special Report 1971.[9]

> Word Knowledge. The Word Knowledge Test is intended as a measure of vocabulary. The "vocabulary" of interest is that which pupils encounter in their ordinary school work ... knowledge of words becomes less exclusively the province of reading instruction as one moves up the grade scale. One begins to want to know how varied the pupil's vocabulary is--not just whether or not he can read words that are in his speaking or listening vocabulary. Thus, the Word Knowledge items give a broad sampling of words from various content areas (science, general experience, travel, etc.) and parts of speech.
>
> Reading. The Reading Test ... is intended as a measure of comprehension of connected discourse presented in a printed medium ... there are four main types of questions which pupils must answer. (1) Pupils must be able to identify the main thought in the selection or select the best title for it. (2) Pupils must be able to identify specific information in the selection or identify the literal meaning of the statements. (3) Pupils must be able to determine from the context the meaning of an unfamiliar word or select the one of several known meanings of a word that is most appropriate in a given context. (4) Pupils must be able to draw inferences from the selection, to identify unstated intentions and moods, and to see relationships between different points in the selection. These inferential-type items receive the greatest emphasis in the test.
>
> Spelling. Close attention was given to the appearance of words in spelling textbooks in selecting words for the Spelling Tests ... words were included in the test to represent a variety of spelling errors ... words are presented in dictated form.

Language. ... coverage is given to punctuation, capitalization, and usage of words. A series of sentences is presented. Different parts of each sentence are underlined. Each underlined part may have one and only one type of error. The error may be in word usage, inadequate punctuation, improper capitalization, or there may be no error at all .... T Elementary Language Test also includes a test of sentence se .... The test of sentence sense is designed to assess the pupil's ability to recognize "telling sentences," "asking sentences," and incomplete sentences.

Mathematics Computation. ... coverage is provided for simple number facts in the basic operations. The test includes horizontal and vertical notation, with more emphasis on the latter ...

Mathematical Concepts. The Mathematics Concepts Test evaluates the pupil's knowledge of fundamental principles and relationships in mathematics.

Mathematical Problem Solving. The Problem Solving Test attempts to evaluate the pupil's total developed ability in mathematics. It demands reasoning with numbers and operations. It presents the pupils with everyday problems in consumer economics, practical measurement ... and other typical situations requiring numerical ability ... (1+) includes problems demanding use of all four fundamental operations and some multiple-step problems. An attempt has been made to keep the vocabulary load and computational load easy ...

Summary

The E-B Direct Instruction Follow Through Model is primarily aimed at teaching basic skills better than typically has occurred in an attempt to catch children from Low Income families up with their middle class peers by the end of third grade. Through the use of carefully programmed lessons, good planning for the use of time, extra "teachers", training, and monitoring of progress, the model seeks to reach its goal. While not explicitly stated as model goals, no procedures are used which might hinder good social and emotional development. We expect our students to be happy as well as competent.

The major evaluation instruments used in the study are the Wide Range Achievement Test, the Metropolitan Achievement Test, and the Slosson Intelligence Test (to get at progress in language development). A summary of information on the reliability and validity of these instruments was provided in this chapter.

CHAPTER 2

DATA COLLECTION AND ANALYSIS PROCEDURES

The strategy which was adopted during the first year of Follow Through (1968-69) was to test each student at entry and each spring thereafter on the WRAT and the SIT. Each of these tests takes about 20 minutes and can be used with five-year olds who cannot read or write. Each is appropriate for a longer-term follow up of students as planned in Follow Through. They provided measures of reading (decoding), arithmetic, spelling, and general language competency (SIT IQ). In the spring of 1972, the MAT was added to the testing schedule at the end of grades 1, 2, and 3. The MAT provided additional measures of reading comprehension, more extensive testing of arithmetic skills, and additional measures of language, spelling, and at the fifth and sixth grades, science.

DATA COLLECTION PROCEDURES

Who Was Tested?

During the first year, there was no pretesting. After that, pretesting occurred for most students. In the spring of 1969 only Washington, D. C. (Nichols Avenue) was not tested because of a conflict with local testing. After that all sites were tested each year that they continued in the study. As noted in Chapter 1, Chicago (Ogden) was dropped from funding in the fall of 1973, Grand Rapids, Michigan could not come to an agreement with the sponsor after the spring of 1972, and Dimmit, Texas dropped Follow Through in February of 1975.

In the spring of 1973, the SIT was given only to students in entering and third grades because of a decision to cut down on testing costs. This decision was reversed the next year. Williamsburg was not given the MAT in 1975 because of a decision to reduce pressures arising from testing in that site. Outside of these exceptions, every available student entering the program through September 1973, was tested in each site, each year.

Testing Procedures

Each year, the testing was guided by a manual of procedures and a checklist set up before hand to provide step-by-step information. A copy of the 1975 manual is attached as Appendix A of this Technical Report. The manual of procedures was built around the test publisher's manuals so as to be consistent with all standardization requirements. Training began at Oregon (or Illinois for 1969 and 1970) with the managers and supervisors who were to serve as the field supervisors for testing. Under the direction of Dr. Becker, a workshop for test supervisors was held each year to go over in detail all procedures for setting up, giving, monitoring, and checking of the testing program.

On site, a similar training program occurred for the local staff assisting in the testing. All teachers were trained to give the MAT and the aides were trained to monitor the students. The local teachers and aides were also trained to give one of these components of the test battery:

1. Individual administration of the SIT.
2. Oral reading and oral math parts of the WRAT (individual testing).
3. Group administered parts of the WRAT.

To prevent errors in subject identification, birthdates, etc., student names were printed on each test at Oregon and placed in manila envelopes by class, ready for distribution. Extra tests were sent to accommodate new students and omissions. The examiner's names and date of testing were filled in on all tests on site. Upon completion of the tests, they were alphabetized within the class and placed back into the manila envelopes. A class roster was provided to record which students had taken the tests, were absent, or had dropped out of the program.

Testing was carried out monitored by the Oregon supervisors. Occasionally, monitoring reports indicated that the testing conditions were not acceptable (e.g., the students were given help by the tester) and the tests were discarded. During the first year all data from one site was discarded because of a failure to follow standardized procedures. Over the years we have identified two classes where evidence of teaching test responses on the WRAT or SIT was detected. With longitudinal data, unwarranted help stands out clearly as non-uniform jumps followed by no progress.

In the early years, tests were scored the day of the testing so that any problems might be noted and corrected on site. All tests were then rescored by data clerks at Oregon. Starting in 1973, MAT and SIT tests have been computer scored at Oregon.

Stanford Research Institute administered MAT tests to a small sample of Follow Through students after 1970. Approximately 500 third graders from four sites were involved (West Iron County, Flint, Brooklyn, Providence). To reduce duplicate testing, copies were made of

the Stanford test responses. A quality control check was used to insure accuracy. Personnel used in the data transferring were hired by the site director and trained locally. When the transferring was completed and passed checking, all tests were mailed to Oregon.

Fifth and Sixth Grade Studies

In addition to the regular spring testing, a study on fifth and sixth graders was conducted in 1975. Approximately 1,400 students were involved in seven sites. Prior Follow Through students and control groups of students with comparable background were tested. In most cases, students in this follow-up study were given the MAT, SIT, and WRAT tests. On the WRAT, both level 1 and 2 were given, although the spelling section was omitted. The spelling section on the MAT was considered sufficient. The Intermediate level MAT test was administered in most cases. A student information form was filled out for all students. Prior Follow Through students were identified by a roster sent out by Oregon. Control groups were sought which were equivalent to the Follow Through students on factors such as income, parent education, ethnic group, and eligibility for the free lunch program. Under the supervision of project managers, local data staff and parent workers located students and filled out the student information forms. Personnel used for this testing were usually the same as those used for the regular spring testing. In larger sites, a few extra testers were brought in from sites that had already finished their testing. In Tupelo the local district gave their fifth and sixth graders the California Achievement Tests, so the MAT was not given. The CAT reading scores were converted to Metropolitan equivalents by using the Anchor Test Study conversion tables.[1]

Other Information From Sites

Student-Information forms and Final-Report-of-Teaching forms were provided to the site data chiefs in the fall and spring, respectively. It was the data chief's job to insure that such forms were returned completed. The Final-Report-of-Teaching contained information on days completed in each Distar program (Reading, Arithmetic, and Language), a record by child of days absent, and a listing of late adds and early drops from the classroom. Bi-weekly reports of teaching and progress on continuous tests were sent to the sponsor by data clerks each two weeks.

Again, to reduce errors, all reports were made on computer printed forms containing student names and identification numbers.

Further Quality Control Checks

Tests and other reports from the sites go through a routine screening for accuracy prior to being placed on IBM cards and fed to our computer file. A checklist of common errors is used for each kind of data. The first time the data go into the computer, logic checks are made where possible (e.g., $IQ = \frac{MA}{CA}$) and data falling outside of preset limits is isolated and verified or corrected by our data clerks. Next, a preliminary report is prepared and sent to the sites with individual student data, classroom summaries, and grade level summaries. The sites and managers are asked to report to us any potential problems. After this the data are placed into the main data file.

METHOD OF DATA ANALYSIS

All children were assigned unique numbers and their data stored on one continuous record. The available records were then coded by grade

and time of data collection to permit grouping together all children who had WRAT, SIT, and Final Reports of Teaching data from the same time periods.

For example, a child who was in the first grade in the fall of 1970, in the first grade in the spring of 1971, and the second grade in the spring of 1972 was assigned a code of 1F70 1S71 1S72. This method of coding permits analysis of the effects of retentions and skipping, since it isolates retentions clearly, e.g., the code 1S71 1S72 is obviously a retention. It also permits us to group children from the same grade levels or cohort together if we wish, even if their pattern of test records are different. Consider these codes:

Child 1	1F69	1S70	2S71
Child 2		1S70	2S71 3S72
Child 3		1S70	1S71
Child 4		1S70	3S72

Each of these children is in the group (called a cohort) starting the Fall of 1969. By a proper combination of their records, a maximum number of children can be studied in an analysis of gain scores. Note that Child 3 is a retention.

Treatment of Retentions

The way in which children retained one or more grades are handled can have an effect on the interpretation of results. In control-group comparisons, retentions do not present a problem unless different retention policies are used for experimental and control groups. In a norm-referenced comparison, placement of retentions with their aremates (keeping them in their starting cohorts) could produce an underestimate

of the program effectiveness. Placing them with their grademates could lead to an overestimate. As it turns out for the data to be presented, the two methods of analysis produce very similar results. We will therefore make the main presentations with retentions grouped with grademates and later show how little the results change when they are placed with agemates.

K-Starting Versus 1st-Starting

The coding procedure also involved a specification of K-starting and 1st-starting children, so that children starting in kindergarten, but first tested at the end of first grade, would not be confused with children starting in first grade. Since there have been numerous changes in K-starting and 1st-starting status because of the increased use of kindergartens. this specification was made individually by site for each data year. This identification allowed an analysis of the effects of the kindergarten-year on the outcome.

"Gains, Singles and Late-Entering" Students

Coded records were divided into "gains" records, "singles" records, and "late entering" records. "Gains" records were students who started the program at the entry level and who had test scores and Final Reports of Teaching for at least two testings. In the beginning cohorts, many "gains" students were not pretested. "Gains" students do not necessarily complete the program. "Singles" records come from students who have only one test in the file. "Late entering" records are from students who did not start the program at the beginning level. These were analyzed separately since they do not represent a full program impact.

The "gains" records were analyzed in three ways:

1. Summary statistics based on all available data. This gives the largest N. This data base is called MAX-N GAINS.
2. Between-year gains based on the same students. This gives an inbetween N. This data base is called BETWEEN-YEAR GAINS.
3. Pretest (kindergarten or first) to posttest (third grade) gains on the same students. This gives the smallest N. This data base is called PRE-POST GAINS.

The later discussions of data for "singles" and "late entries" will rely on the data presented in our 1973 Technical Report.[2] These earlier analyses include 3,600 of the present 4,883 records in the "singles" file, and 1,020 of the 1,926 in the late entering file. Because this is low-yield information, showing essentially the results previously reported, the data will be briefly described, but not presented.

Standard computer programs were used to compute means, standard deviations, and other summary statistics used to describe the data.

SUBJECTS

All students entering the University of Oregon Engelmann-Becker classrooms from September 1968 through September 1973 on whom we have acceptable data records, are included in this study. There are 21,485 records in the master file. Analysis of these records is shown in Table 2.1.

- - - - -

Table 2.1

- - - - -

Table 2.1

Breakdown of Records in the Computer File

RECORDS EXCLUDED FROM ANALYSIS			929
Names but no data	536		
Bad Codes	6		
Head Start Planned Variation Only	140		
Class entered above program (used as within site control groups)	247		
"SINGLES" ANALYSIS			4,883
"LATE ENTERING" ANALYSIS			1,826
	All Kids	Low-Income Kids	
K-starting	1,122	749	
1st-starting	704	515	
"GAINS" ANALYSIS			13,847
	All Kids	Low-Income Kids	
K-starting	6,995	5,922*	
1st-starting	6,855	5,565*	
PRE-POST GAINS ANALYSIS			
	All Kids	Low-Income Kids	
K-starting	435	374*	
1st-starting	1,208	1,080*	
TOTAL RECORDS			21,485

*Actual analyses will always have smaller N's because of partially missing data.

Excluded Records

Names but no data are most likely from children who entered the program without being pretested and left before the first posttest. Bad codes are records where an error in grade or date makes the code impossible, and where we have not been able to correct the error. The Head Start Only children participated in a sponsored Planned Variation Head Start but did not go on to Follow Through. Class entered above program students are students who were included in Follow Through but did not go through the model program in kindergarten or first grade.

Retentions and Skips

In K-starting sites, there have been 350 retentions for one year (6.2% of Low-Income students in "gains" file), 9 retentions for two years, and 17 skips of one year. In 1st-starting sites, there have been 606 retentions for one year (11% of Low-Income students in "gains" file), 22 retentions for two years, and two skips of one year.

Breakdown of Students by Site

Tables 2.2 to 2.3 give the number of students in the "gains" analysis by site and cohort (retentions and skips with grademate). Cohort 1 started in the fall of 1968; cohort 6 started in the fall of 1973. A breakdown and discussion of the "singles" and "late entering" groups by site and cohort can be found in our 1973 Technical Report.[3]

- - - - - - - -

Table 2.2 - 2.3

- - - - - - - -

Tables 2.2 and 2.3 permit the reader to see the constitution of the data base in terms of cohorts (starting years) and sites. The comparison

Table 2.2

N for All and Low-Income K-Starting Students in the MAX-N-GAINS Analysis by Site and Cohort. Retentions with Grademates

K-Starting Sites	Cohorts													
	1		2		3		4		5		6		Total	
	All	Low Income	All	Low Income	All	Low Income	All	Low Income	All	Low Income	All	Low Income	All	Low Income
Rosebud Cap	-	-	36	34	73	64	76	68	55	30	86	42	326	238
Flippin	-	-	25	23	24	24	35	32	52	52	39	34	175	165
Cherokee	-	-	-	-	100	84	118	105	100	81	97	72	415	342
Grand Rapids	187	157	232	221	248	242	267	265	8	0	-	-	942	885
Racine	90	60	122	96	141	109	134	93	161	146	201*	50	849	554
West Iron Co.	53	31	60	46	40	36	44	39	24	23	28	21	249	196
Flint	-	-	130	82	167	95	91	69	100	89	115	89	603	424
Todd County	1**	1	117	90	135	118	169	153	138	130	145	120	705	612
Chicago	-	-	32	12	36	15	45	20	24	10	-	-	137	57
Washington	-	-	46	45	82	80	73	65	95	89	101	70	397	349
Brooklyn	44	37	56	56	54	54	54	54	90	89	87	78	385	368
Providence	1	1	117	117	224	218	186	182	185	185	203	189	916	892
E. St. Louis	-	-	-	-	116	80	237	227	280	276	263	257	896	840
Totals	376	287	973	822	1440	1219	1529	1372	1312	1200	1365	1022	6995	5922
TOTAL														

*Most of these must be Low Income. This is a miscoding that has to be corrected yet.

**These small groups arise from placing retentions and skips with grademates.

Table 2.3

N for All and Low-Income 1st-Starting Students in the MAX-N-GAINS Analysis by Site and Cohort. Retentions with Grademates.

1st-Starting	Cohorts													
	1		2		3		4		5		6		Total	
	All	Low Income	All	Low Income	All	Low Income	All	Low Income	All	Low Income	All	Low Income	All	Low Income
Rosebud Cap	30	29	86	70	-	-	-	-	-	-	-	-	116	99
Dimmitt	-	-	173	86	137	66	142	82	172	97	190	133	814	464
Flippin	-	-	26	23	1*	1	-	-	-	-	-	-	27	24
Smithville	28	28	80	59	76	66	85	70	81	50	90	67	440	340
Tupelo	86	83	109	99	95	91	117	108	99	86	105	96	611	563
Cherokee	-	-	-	-	110	103	16	16	-	-	-	-	126	119
Racine	68	35	10	9	-	-	-	-	-	-	-	-	78	44
E. Las Vegas	115	89	102	86	105	95	110	93	128	101	112	96	672	560
Uvalde	115	88	147	116	161	133	135	111	126	124	119	118	803	690
Todd County	92	85	133	122	20	18	3	3	1	1	-	-	249	229
Dayton	266	194	221	166	252	181	243	181	256	180	302	226	1540	1128
Chicago	-	-	30	13	2	2	-	-	-	-	-	-	32	15
Williamsburg	-	-	130	130	143	143	102	100	145	142	156	122	676	637
Providence	-	-	131	131	8	8	-	-	-	-	-	-	139	139
E. St. Louis	202	195	147	140	178	175	5	4	-	-	-	-	532	514
Totals	1002	826	1525	1250	1288	1082	958	768	1008	781	1074	858	6855	5565

TOTAL All K and 1st Starting - 13,847

*These small groups arise from placing retentions and skips with grademates.

of the N's for All and Low Income show that 80 percent of the students in our Follow Through classrooms come from Low-Income families as defined by the federal poverty guidelines.

The racial composition of the sites is as follows:

Mostly Native American:	Todd County, S. Dakota Rosebud Tribe, S. Dakota Cherokee, N. Carolina
Mexican American:	Dimmitt, Texas (50 percent) Uvalde, Texas (90 percent)
Spanish:	E. Las Vegas, N.Mexico
Mostly Black:	Tupelo, Mississippi Grand Rapids, Michigan Flint, Michigan Dayton, Ohio Washington, D. C. Williamsburg County, S. Carolina Brooklyn, N. York E. St. Louis, Illinois
Mixed White and Black:	Racine, Wisconsin Chicago, Illinois Providence, Rhode Island
Mostly White:	Flippin, Arkansas Smithville, Tennessee West Iron County, Michigan

The N's in Table 2.2 and 2.3 are usually larger than the N's presented in later analyses because some children do not have all tests each year.

INTERPRETATION OF DATA PRESENTATIONS

When presenting data which are referenced to a national norm population, we will provide percentiles, mean standard scores and their standard deviations, and mean grade-equivalent scores. The following abbreviations are used:

N = Number of cases
%-tile = Percentile for mean standard score
S. S. = Mean standard score
S. D. = Standard deviation of the standard scores
G. E. = Mean grade-equivalent score

In referring to tests and test variables the following abbreviations will be used in the tables:

WRAT = Wide Range Achievement Test

R = Reading (decoding)
A = Arithmetic
S = Spelling

MAT = Metropolitan Achievement Test

TR = Total Reading

WK = Word Knowledge
WA = Word Analysis
R = Reading (comprehension)

TM = Total Math

MCom = Math Computation
MCon = Math Concepts
MPS = Math Problem Solving

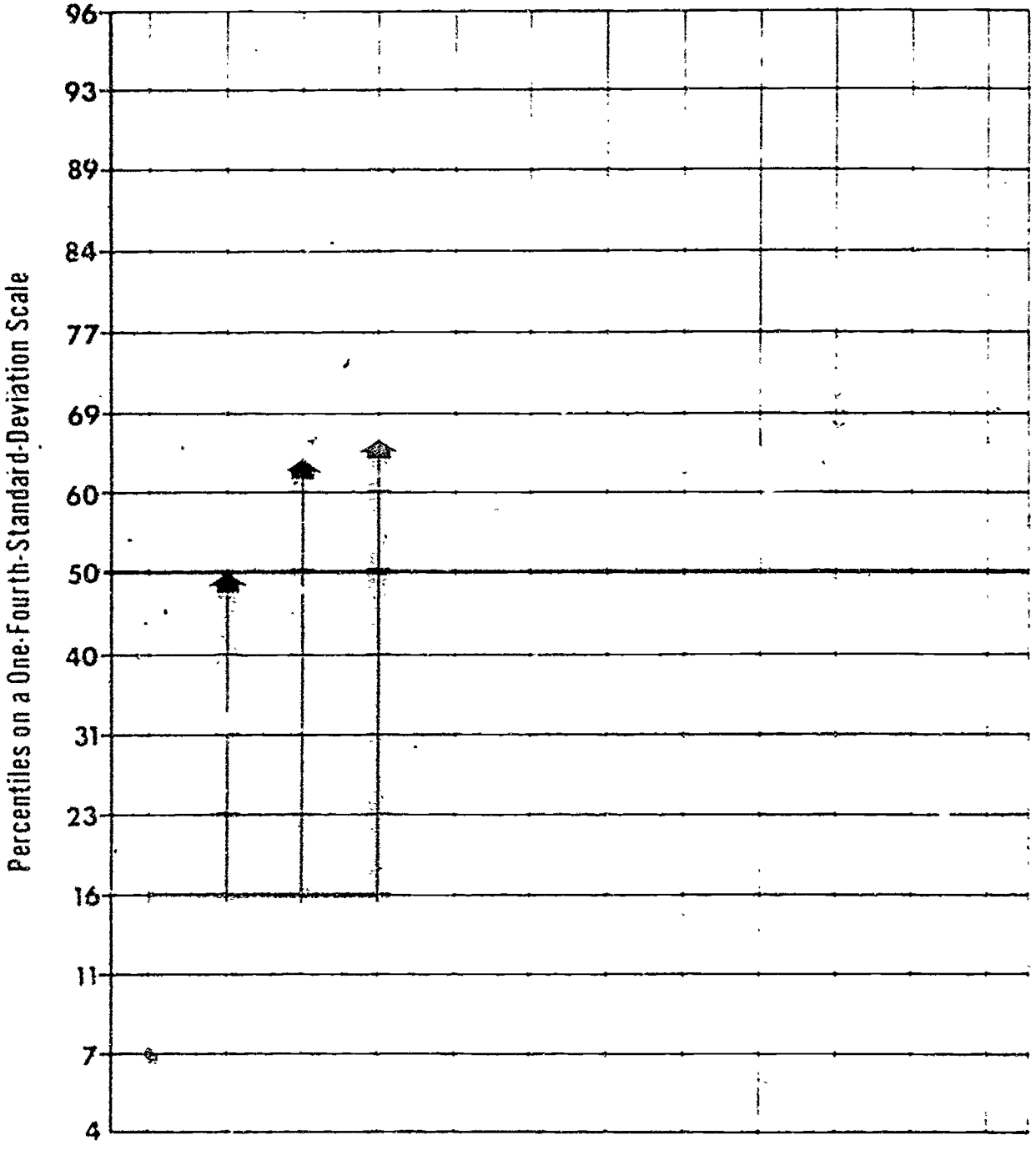
Sp. = Spelling

Lan. = Language

In graphing norm-referenced comparisons we will present the data in percentiles on a one-fourth standard deviation scale. Since percentiles are easily interpreted with 50 being the median for the norm group, they are preferred in presenting results to most groups. However, percentile changes near 50 tend to overemphasize differences, while those near the extremes tend to underemphasize differences. This problem can be corrected for by plotting percentiles on a standard-deviation based scale. Furthermore, when this is done, the sizes of differences shown can be directly interpreted for their educational significance. Generally, a one-fourth standard deviation gain has been accepted as a minimum criterion for educational significance.[4,5] Table 2.4 illustrates a norm-referenced

Table 2.4

Hypothetical WRAT Reading Percentiles by Grade Level



PRE	
POST	
N	
%-tile	
S S	
S. D.	
G. E	

comparison for hypothetical WRAT reading data. The percentile for the mean standard score at pretest is 16. This is one standard deviation below the national mean. This pretest average is then taken as a reference point for improvement on the national norms at the end of first (1), second (2), and third (3) grades. The size of the improvement is shown by the length of the arrow in one-fourth standard deviation units.

The reader should carefully note the differences in the nature of standard scores and percentiles used for the WRAT and the MAT. For the WRAT, standard scores are defined for a given age group and have a mean of 100 and a standard deviation of 15. We have computed our summary statistics from these standard scores (as well as grade equivalents) and then converted the mean standard score to a percentile. This conversion assumes a normal distribution of the standard scores in the norm group (a reasonable assumption).

The MAT uses expanded standard scores to provide a basis for comparison of scores on the same measure from different test levels and different forms. The mean standard score increases an average of 10 points a year and has a standard deviation of 10. These scores make it possible to directly compare standard scores from the Primary I test with those from the Elementary level test, etc. To get percentiles, raw scores were converted to standard scores. Then statistics were computed on the standard scores and the mean standard scores were converted to percentiles.

The grade-equivalent scores reported are the means of the individual grade-equivalent scores unless otherwise specified. These

means will tend to be higher than the grade equivalents of the median score (50th percentile). For example, on MAT Total Reading, the grade equivalent of the median is 3.5 at the end of third grade rather than 3.8.

Summary

In this chapter we have described the procedures used to collect the data on our students. We have also detailed the methods of data analysis used to enable us to focus the analysis on full-term Follow Through children from Low-Income families. We have taken special care to note two ways of dealing with retentions and skips, placing them with agemates and grademates. For our primary presentations they will be placed with grademates. We have defined the basis for three kinds of "gains" analyses, the MAX-N GAINS, the BETWEEN-YEARS GAINS, and the PRE-POST GAINS. Because of sample limitations at this time, the PRE-POST GAINS analysis will be useful primarily with 1st-starting sites.

The composition of the students in the "gains" analysis was detailed by site, by cohort, by income status, and ethnic group.

Finally, a guide to interpretation of the data presentations was provided. We are now ready to get into the actual findings.

CHAPTER 3

RESULTS FOR KINDERGARTEN-STARTING SITES AND FIRST-GRADE-STARTING SITES

In this chapter we will address ourselves to the overall evidence pertaining to objectives 1 and 2, namely:

1. To demonstrate that economically disadvantaged children can be taught so that they will show a rate of progress typical for national norm groups in reading and math for grades K to 3.

2. To demonstrate economically disadvantaged children can be taught so that they show a mean IQ gain from entry grade through third grade, even though the usual expectancy is for a loss on IQ tests.

ACHIEVEMENT TEST DATA FROM THE MAX-N GAINS ANALYSIS

The major presentation of findings will use the data from the MAX-N GAINS analysis as described in the last Chapter. Later in this chapter we will present a comparison of the various gains analyses to show that the MAX-N GAIN analysis provides a conservative picture of results while utilizing the largest data base.

K-Starting Low-Income Students

A major goal of the E-B model has been to get Low-Income Follow Through students at a par with national norms by the end of third grade. Table 3.1 gives one overall indication of the degree to which we have been successful in meeting this objective. Using the entry level performances on the WRAT in Reading, Arithmetic, and Spelling as the baselines, subtotal gains are apparent against the national norm group on all measures. The length of the arrows expresses the magnitude of gains against the norms in quarter standard deviation units. The

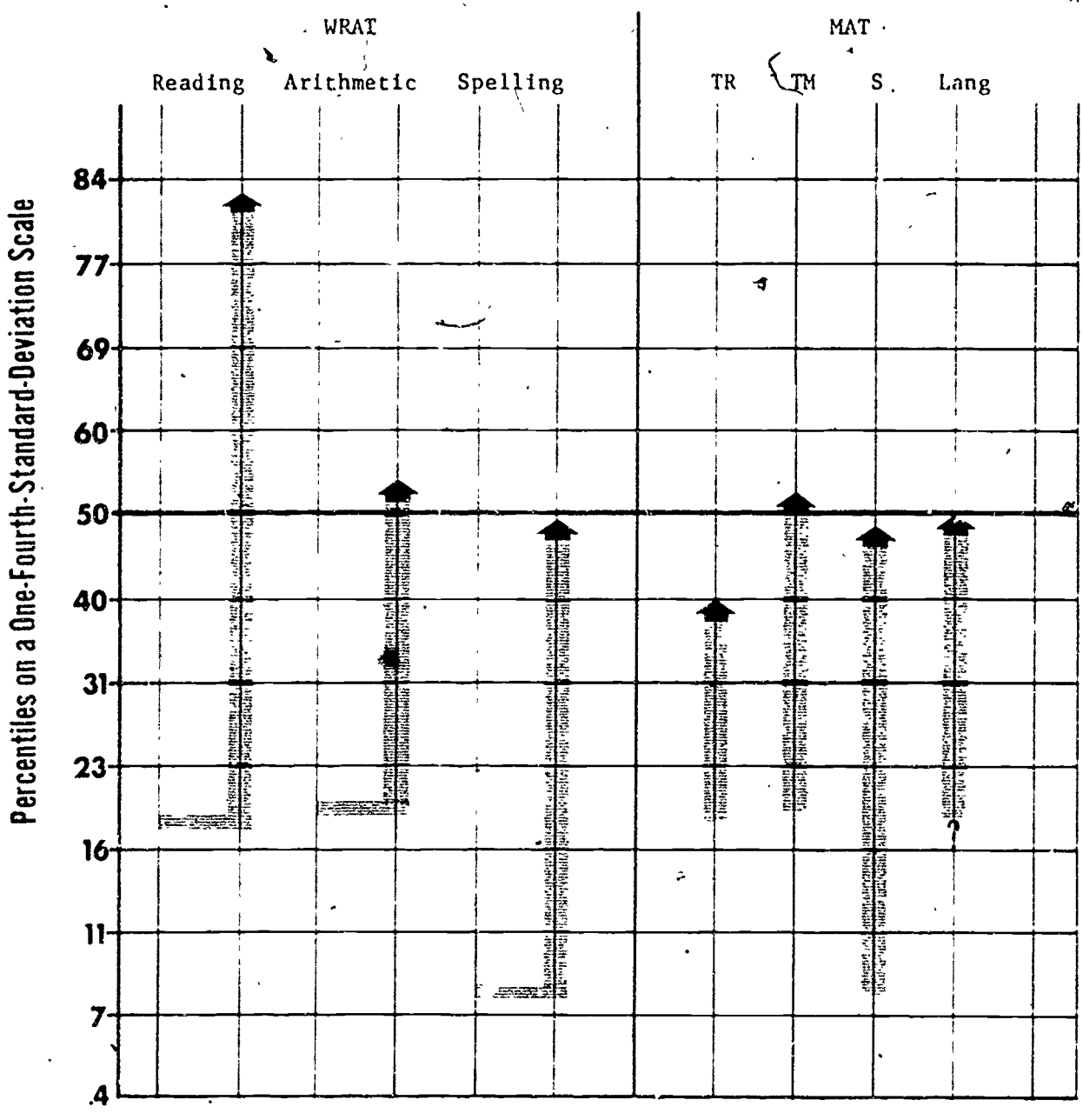
- - - - - - - -

Table 3.1 here

- - - - - - - -

Table 3.1

Norm-Referenced Gains on the WRAT and MAT from Kindergarten through Third Grade for K-Starting, Low-Income, E-B Follow Through Students



PRE	K		K		K						
POST		3		3		3	3	3	3	3	
N	2435	1988	2676	1989	2520	1974	1863	1804	1824	1854	
%-tile	18	33	19	54	8	49	40	52	48	50	
S. S.	85.9	114.5	86.6	101.2	79.5	99.7	58.0	71.1	62.8	69.9	
S.D.	16.2	20.9	15.8	8.6	20.4	13.9	10.3	11.9	11.3	13.2	
G. E.	.18	5.35	.19	3.99	-.09	3.79	3.31	3.98	3.68	4.28	

children have learned decoding skills well (WRAT Reading) and fall above or very near the 50th percentile (national median) on all measures except MAT Total Reading, where they fall short by one-quarter standard deviation (40th percentile).

A better perspective of the meaning of these findings is gained by comparing the MAT performances of the E-B students as shown in Table 3.1 with that of the students in the Follow Through National Evaluation of Cohort II.[1] The Cohort II study compared Follow Through (FT) students starting kindergarten in 1970 with Non-Follow Through (NFT) comparison groups. The Follow Through students were taken from 13 different sponsors including E-B. Table 3.2 shows the E-B MAT data in comparison to the overall end of third grade findings for FT and NFT groups in Cohort II. The performance of the E-B students is higher in each case. In comparison to 13 Follow Through sponsors, our MAT scores are one-half standard deviation higher in Reading, a full standard deviation higher in Math, and a half standard deviation higher in Spelling. We lack the FT comparison figure for Language.

The Non-Follow Through control group has been found by ABT to be superior to the Follow Through group on most measures of socioeconomic status. For example, their average income was $5964 versus $4733 for Follow Through, there were fewer females as household heads (34% versus 43%) and more Whites (34% versus 19%). The common experience is that the most needy were designated for Follow Through and the control groups were selected from what was left. Even given these advantages, it can be seen in Table 3.2 that the E-B children outperform the NFT group on every measure. The performances E-B students on MAT Math and Language are particularly impressive.

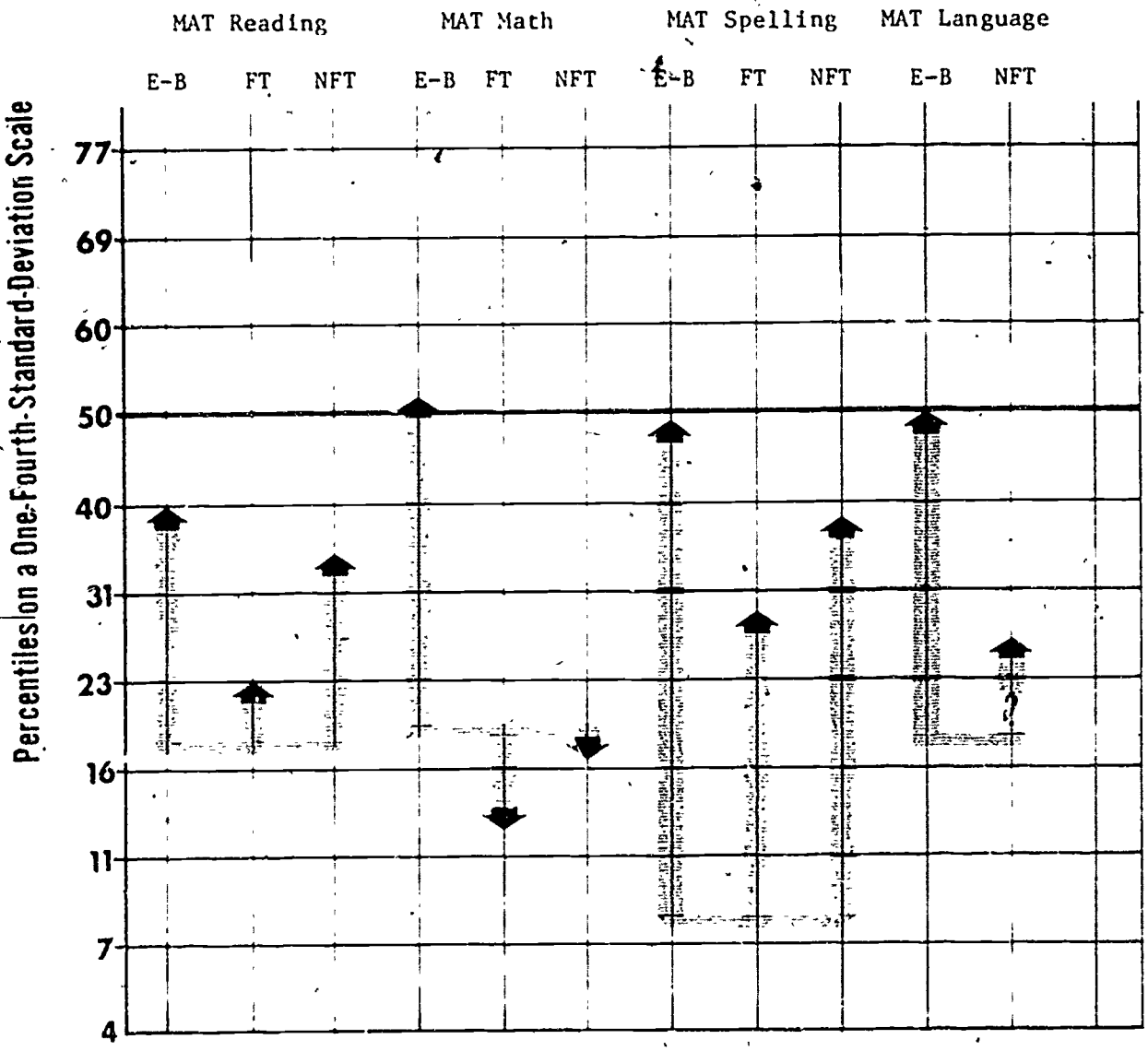
- - - - - - -

Table 3.2 here

- - - - - - -

Table 3.2

MAT scores at the end of third grade for K-starting Low-Income students in the E-B program, in programs sponsored by 13 Follow Through Sponsors (FT), and in the control groups for the 13 Follow Through Sponsors (NFT). FT and NFT data are from the National Evaluation of students entering kindergarten in 1970.[1]



	MAT Reading			MAT Math			MAT Spelling			MAT Language	
	E-B	FT	NFT	E-B	FT	NFT	E-B	FT	NFT	E-B	NFT
POST	3	3	3	3	3	3	3	3	3	3	3
N	1863	3367	2151	1804	3367	2151	1824	3367	2151	1854	2151
%-tile	40	23	34	52	13	17	48	28	38	50	27
S. S.	58.0	52.4	56.0	71.1	57.4	59.3	62.8	56.4	60.0	69.9	60.4
S. D.	10.3	--	--	11.9	--	--	11.3	--	--	13.2	--
Mean G. E.	3.31	--	--	3.98	--	--	3.68	--	--	4.28	--
Mean S.S. → G.E.	3.2	2.6	3.0	3.8	2.8	2.9	3.5	2.8	3.1	4.0	3.0
%-tile G.E.	(3.5)			(3.75)		(3.1)	(3.6)			(4.0)	

We have placed a number of types of grade-equivalent scores at the bottom of Table 3.2 to illustrate some of the problems in their interpretation. The Mean G. E. which is what we will typically report, will usually give a higher score than that obtained by finding the grade equivalent for a mean standard score (Mean S. S. → G. E.). This is because there are more extreme grade equivalents above the median than below the median. Also note that the grade equivalent score for the 50th percentile (50%-tile G. E.) varies greatly from scale to scale on the MAT (as much as a half a grade level) Only by comparing the mean standard score grade equivalent with the 50th percentile-grade-equivalent can one fairly judge how good a performance is relative to the norms when using grade equivalents. The direct plotting of the percentile values for mean standard scores, as we will usually do in this report, circumvents these problems.

Outcomes by Grade Level

Table 3.3 shows the WRAT Reading and Arithmetic performances for E-B Low-Income students by each grade level. Note that the N's for some grade levels exceed 5000. The pretesting N is low because the first cohort was not tested at all on pretest and others were only partially tested. The third grade N is down because two cohorts have not yet reached third grade. The by-grade analysis for reading shows that gains relative to the norm groupwere made in kindergarten and first grade on WRAT Reading and then maintained thereafter. The data for arithmetic show that there is a major gain from kindergarten in-

struction which is then partly lost, but a better than average performance is maintained. The analysis by cohorts (to be presented later) shows a subtantial improvement by later cohorts.

- - - - - - -

Table 3.3

- - - - - - -

Table 3.4 presents the MAT Reading, Math, and Spelling scores by grade level. The MAT Total Reading scores show a decline of three-fourths of a standard deviation from the end of first grade to the end of third grade in comparison to the norm group. Such a decline is not found for Math or Spelling.

- - - - - - -

Table 3.4

- - - - - - -

Table 3.5 shows that the decline in standard scores for MAT Reading is consistent for all three sub-scores; Word Knowledge, Word Analysis, and Reading. An attempt to interpret this finding will be presented in the Discussion Chapter.

- - - - - - -

Table 3.5

- - - - - - -

The analysis of MAT Math subtests show inconsistent minor variations with grade level. The trend is toward some continued improvement against the norm group from second to third grade. Our students do especially well in math computation at the end of third grade.

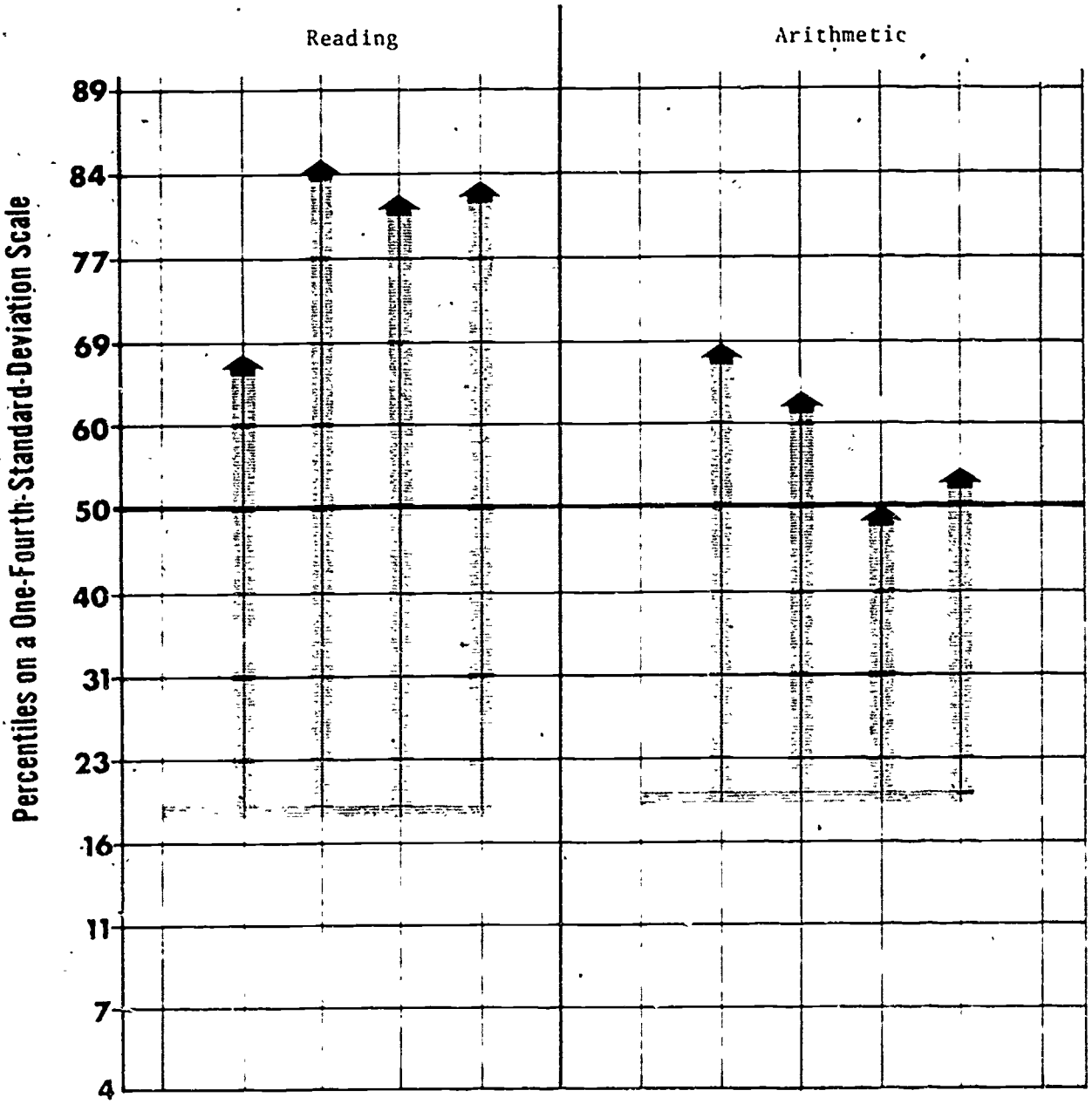
- - - - - - -

Table 3.6

- - - - - - -

Table 3.3

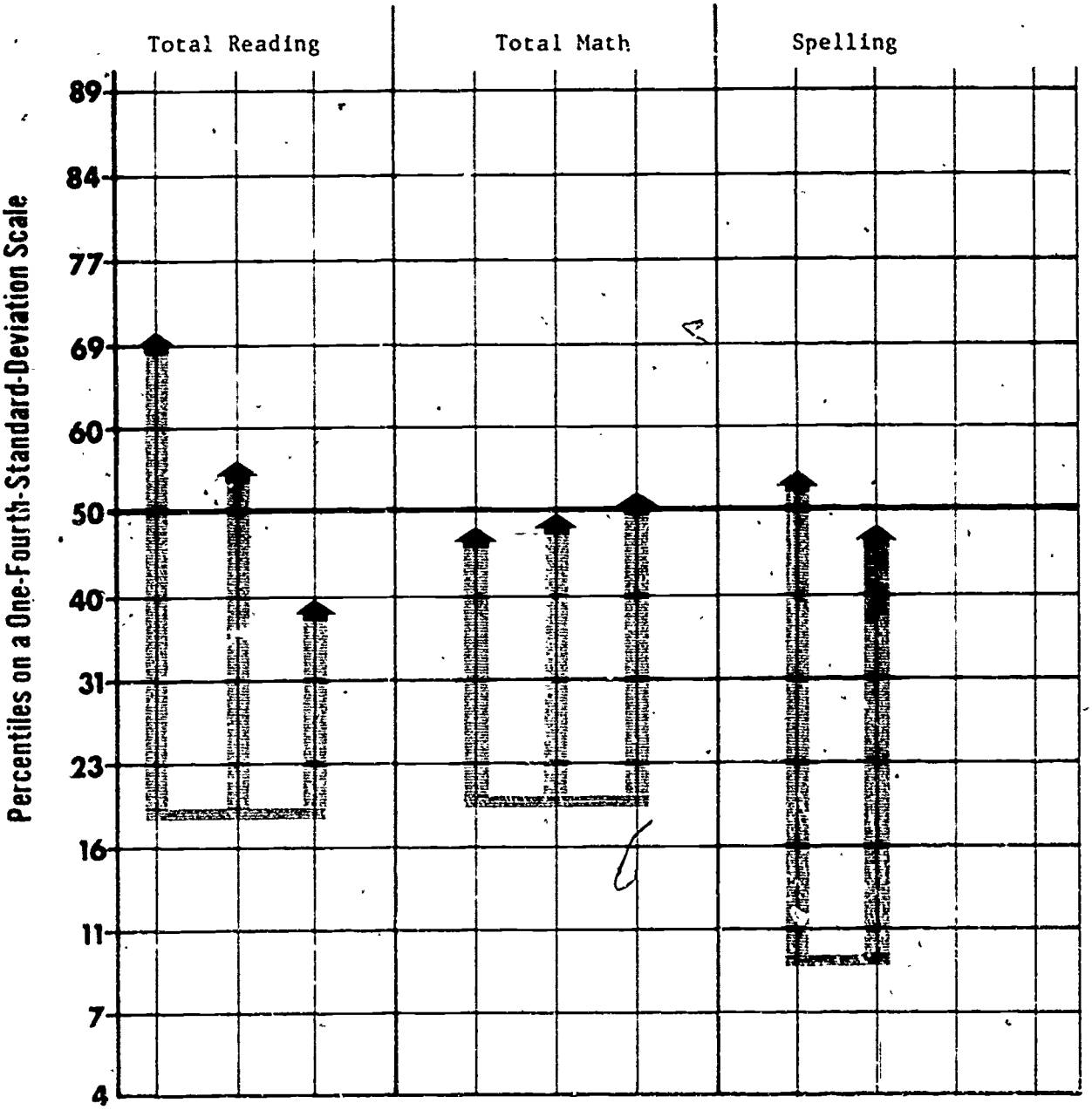
Norm-Referenced Gains on the WRAT by Grade Level for K-Starting, Low-Income, E-B Follow Through Students



PRE	K					F				
POST		1	2	3		K			3	
N	2435	5181	4819	3309	1985	2676	5189	4809	3290	1989
%-tile	18	68	85	82	93	29	69	64	56	54
S. S.	85.9	107.0	113.7	113.0	114	88.6	107.8	106	106.0	106
S. D.	16.7	18.7	20.3	19.8	20.9	15.8	14.6	14.1	14.9	15
G. E.	.18	1.34	2.41	4.23	6.2	.19	1.38	2.38	4.21	6.00

Table 3.4

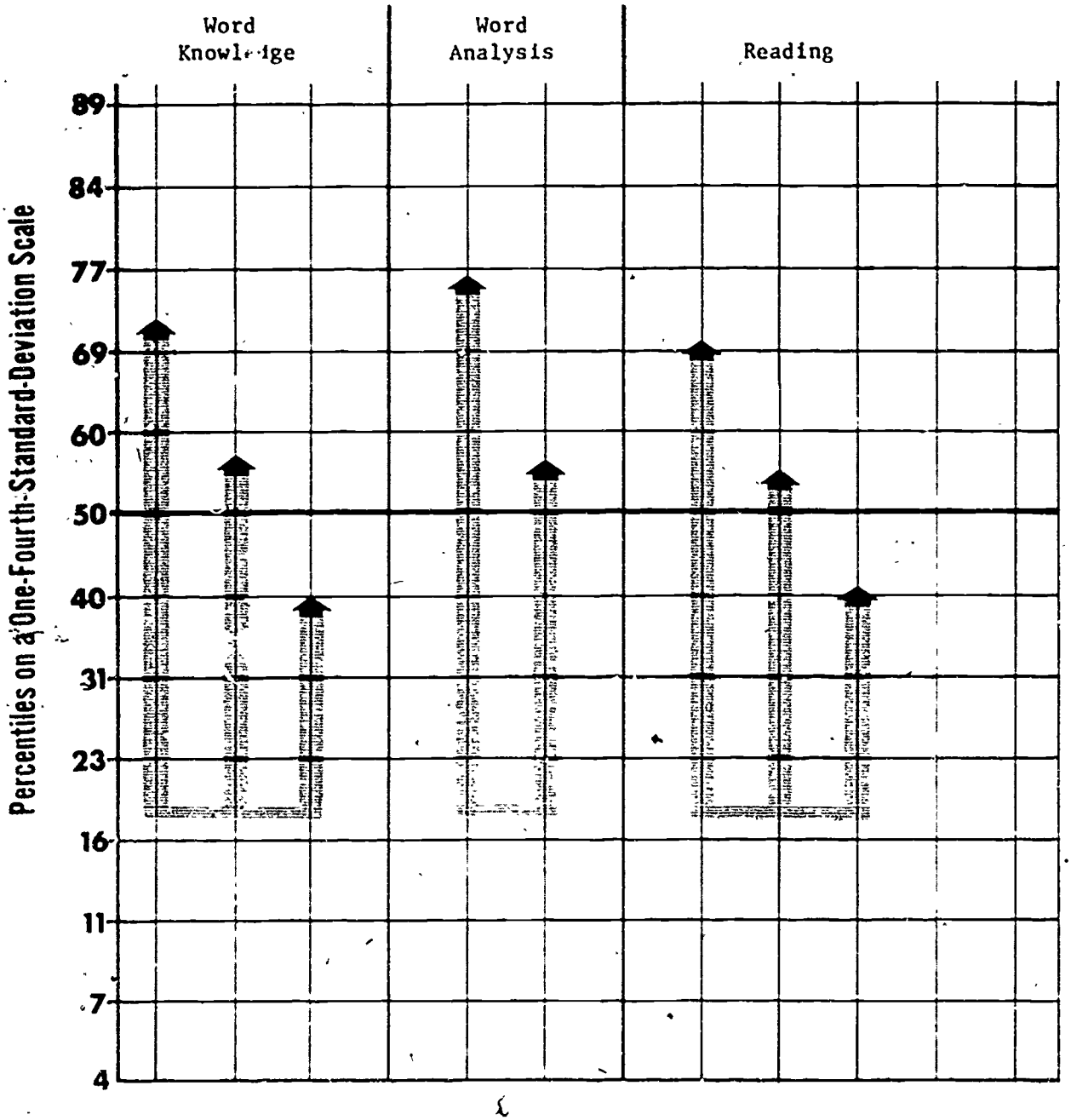
Norm-Referenced Gains on the MAT by Grade Level for K-Starting, Low-Income E-B Follow Through Students



POST	1	2	3	1	2	3	2	3
N	2738	2289	1863	2761	2305	1804	2275	1785
%-tile	70	56	40	47	50	52	54	47
S. S.	44.0	54.0	58.0	40.7	57.0	71.1	57.4	62.8
S. D.	9.9	8.5	10.3	11.0	10.4	11.9	10.6	11.3
G. E.	2.19	2.89	3.31	1.82	2.78	3.98	3.08	3.68

Table 3.5

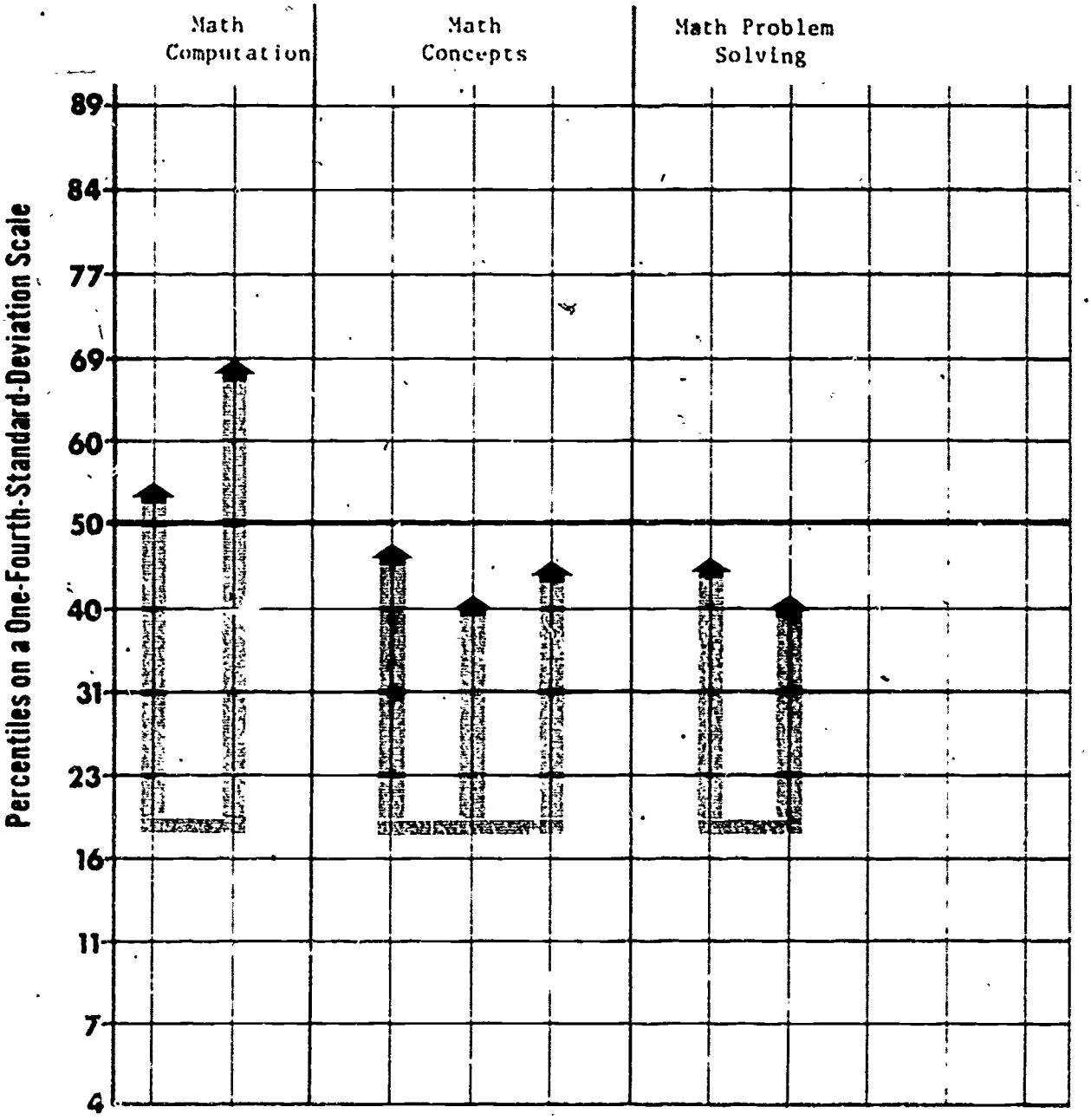
Norm-Referenced Gains on MAT Reading Subscales by Grade
for K-Starting, Low-Income, E-B Follow Through Students



	Word Knowledge			Word Analysis		Reading		
PRE								
POST	1	2	3	1	2	1	2	3
N	2750	2306	1867	2745	2341	2757	2306	1865
%-tile	72	57	40	76	56	70	55	41
S. S.	46.4	55.3	59.9	42.9	53.5	44.0	54.2	58.4
S. D.	11.3	8.4	9.6	7.7	8.5	10.8	9.1	11.1
G. E.	2.31	2.94	3.42	2.15	3.08	2.13	2.85	3.29

Table 3.6

Norm-Referenced Gains on MAT Math Subscales by Grade
for K-Starting, Low-Income, E-B Follow Through Students



POST	2	3	1	2	3	2	3
N	2334	1824	2761	2336	1820	2335	1820
%-tile	55	69	47	42	45	46	42
S. S.	54.5	70.6	40.7	54.0	65.5	56.0	65.6
S. D.	10.5	11.9	11.0	11.1	12.1	11.8	12.6
G. E.	2.93	4.39	1.82	2.81	3.89	2.91	3.74

1st-Starting Low-Income Students

In 1st-starting sites, there is one less year to teach the children. Also several of our 1st-starting sites in the Deep South and Southwest have our lowest performing children on entry in terms of basic language competencies. Even given these limitations, the data in Table 3.7 show that Low-Income students from 1st-starting sites are at or above the national norms at the end of third grade on WRAT Reading, MAT Math, and MAT Language. On both the WRAT and the MAT they fall one-quarter standard deviation short of the norm on Spelling, and one-half standard deviation short of MAT Reading.

- - - - - -

Table 3.7

- - - - - -

Outcomes by Grade Level

Table 3.8 shows a progressive improvement against the national norms over grade levels on WRAT Reading. In Arithmetic and Spelling, the only change with grade levels is the finding (common for our students) that they do not score well on WRAT Arithmetic during second grade.

- - - - - -

Table 3.8

- - - - - -

The findings over grade levels for the MAT show an improvement against the norms from second to third grade in Math, and losses in Reading and Spelling.

- - - - -

Table 3.9

- - - - -

Table 3.7

Norm-Referenced Gains on the WRAT and MAT through Third Grade for 1st-Starting, Low-Income, E-B Follow Through Students

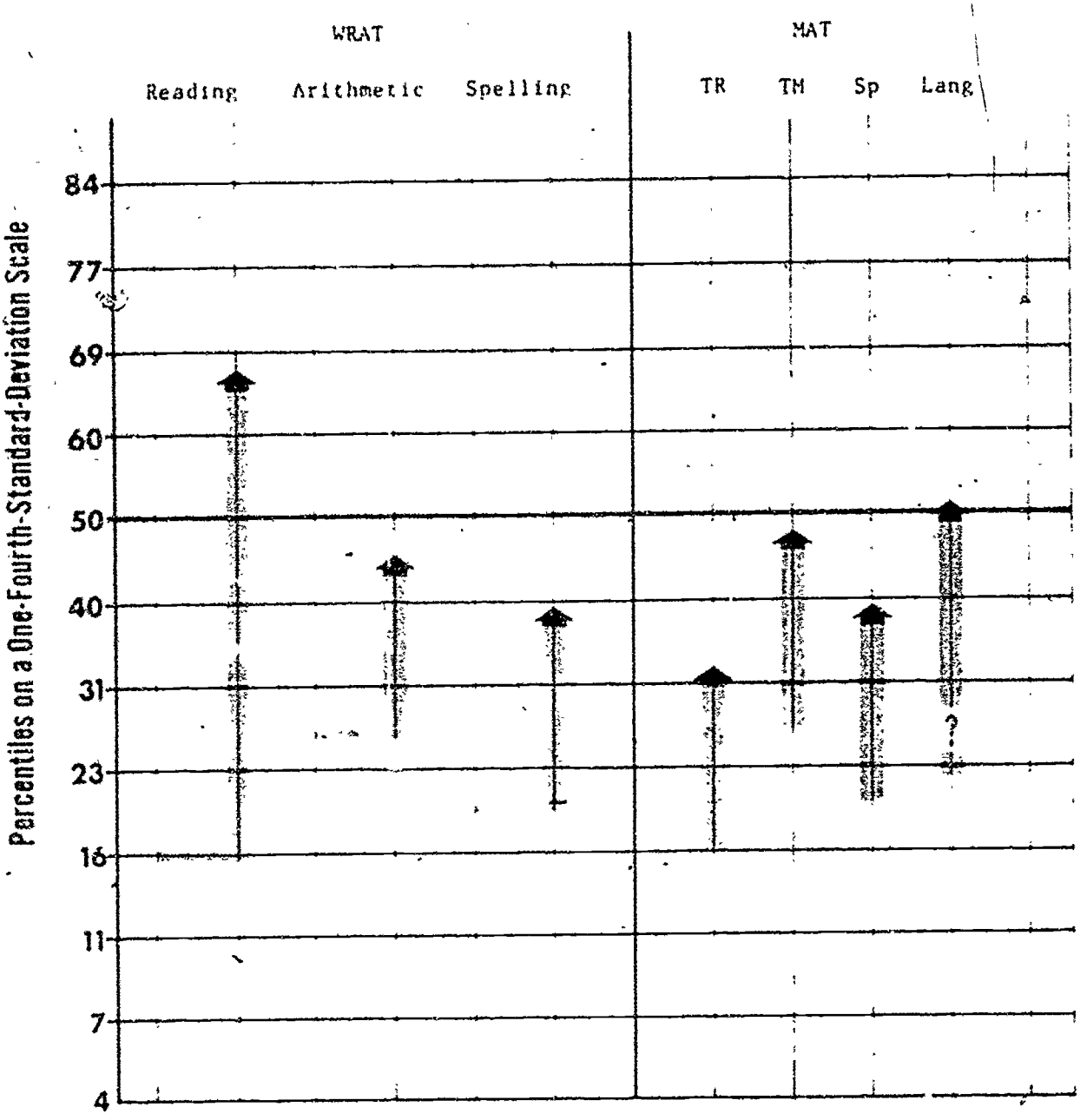
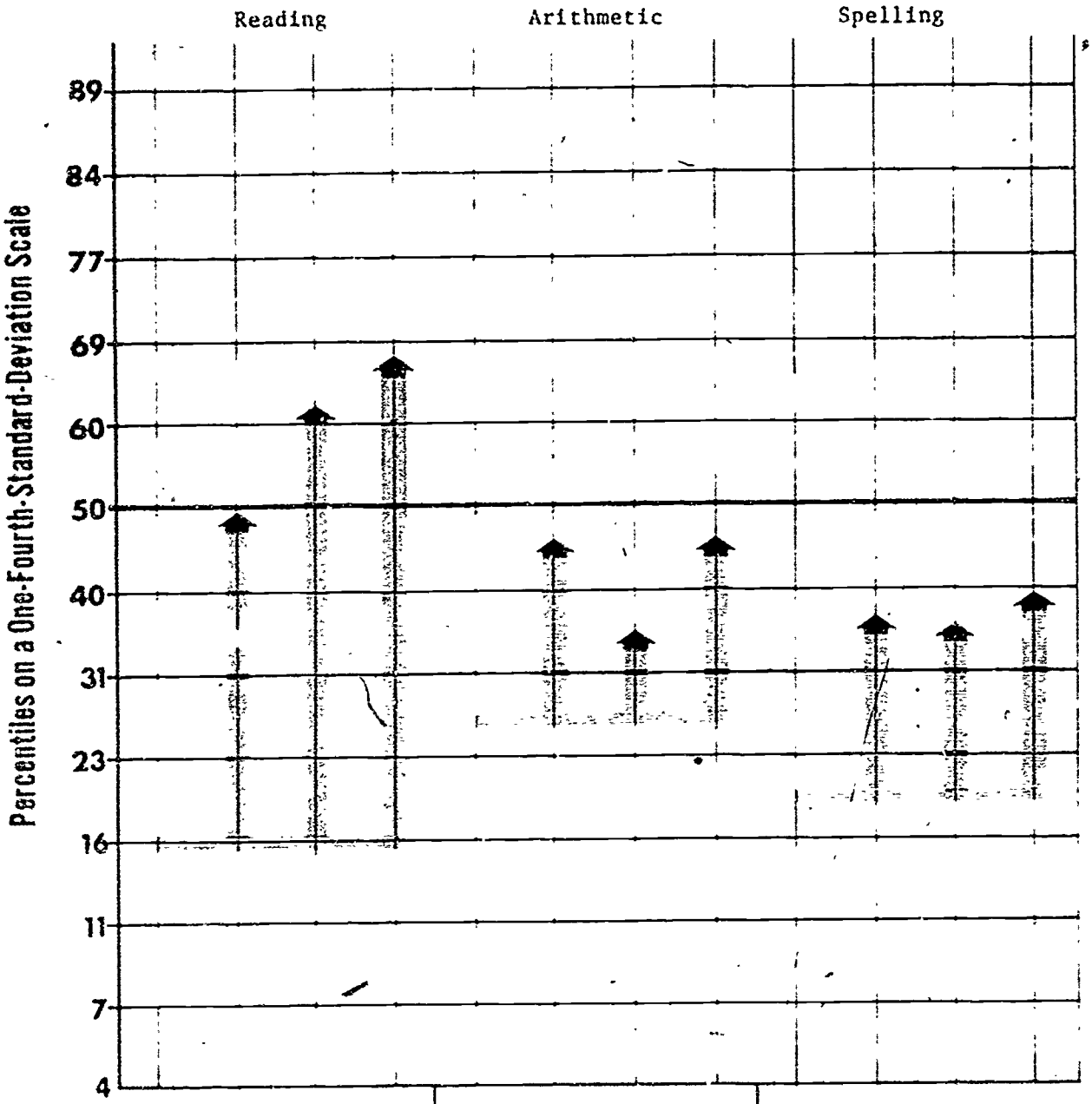


Table 3.8

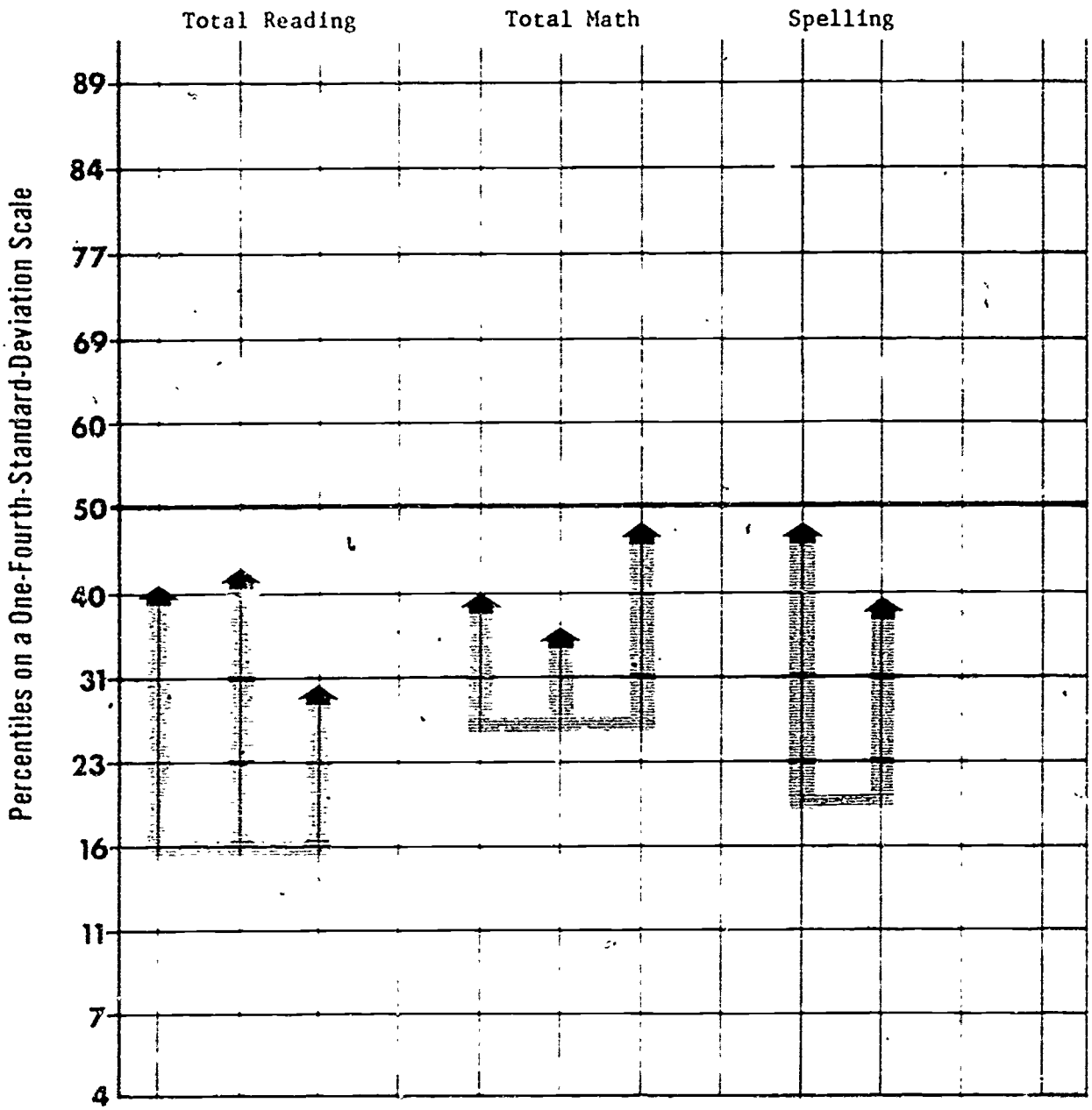
Norm-Referenced Gains on the WRAT by Grade Level
for 1st-Starting, Low-Income, E-B Follow Through Students



	Reading				Arithmetic				Spelling			
PRE	1				1				1			
POST		1	2	3		1	2	3		1	2	3
N	2412	5160	4665	3629	2444	5136	4661	3630	2421	5173	4646	3609
%-tile	16	49	63	67	26	45	35	45	20	37	36	39
S. S.	84.9	99.6	104.8	106.5	90.4	97.9	94.7	97.9	87.4	95.1	94.5	95.9
S. D.	14.4	18.2	18.4	19.2	13.6	12.1	9.0	9.8	15.1	12.9	13.1	15.1
G. E.	.51	2.02	3.54	4.66	.77	1.91	2.69	3.78	.64	1.75	2.69	3.59

Table 3.9

Norm-Referenced Gains on the MAT by Grade Level
for 1st-Starting, Low-Income, E-B Follow Through Students



POST	1	2	3	1	2	3	2	3
N	1266	1681	1931	1269	1676	1926	1654	1863
%-tile	41	43	30	40	35	48	48	39
S. S.	37.4	50.3	55.0	38.0	54.3	69.9	56.2	60.5
S. D.	8.6	7.4	8.9	10.4	8.6	11.1	10.2	10.8
G. E.	1.81	2.60	2.99	1.68	2.58	3.84	2.96	3.41

Analysis of MAT Reading and Math subscales over grade levels (Tables 3.10 and 3.11) show a superior performance of our 1st-starting students on Word Analysis in first grade (relative to the other reading measures), and a superior performance on Math Computation in third grade. These findings are consistent with known program strengths.

- - - - - - - - - - -

Table 3.10 and 3.11

- - - - - - - - - - -

Comparison of Results for K-Sites and 1st-Sites

Table 3.12 compares the mean standard scores on MAT and WRAT tests for K-starting and 1st-starting sites. The entry points on the WRAT are generally comparable or a little lower for K-starting children. By the end of third grade, there is a clear and highly significant advantage for the K-starting students. On WRAT Reading the advantage is more than one-half standard deviation. On WRAT Arithmetic and Spelling, the advantage is one-quarter standard deviation. On the MAT, the advantage is about one-quarter standard deviation for each measure except Language (which is not shown n Table 3.12). There is no difference on the Language subtest. This test measures grammatical skills which are taught to all our children in second and third grades.

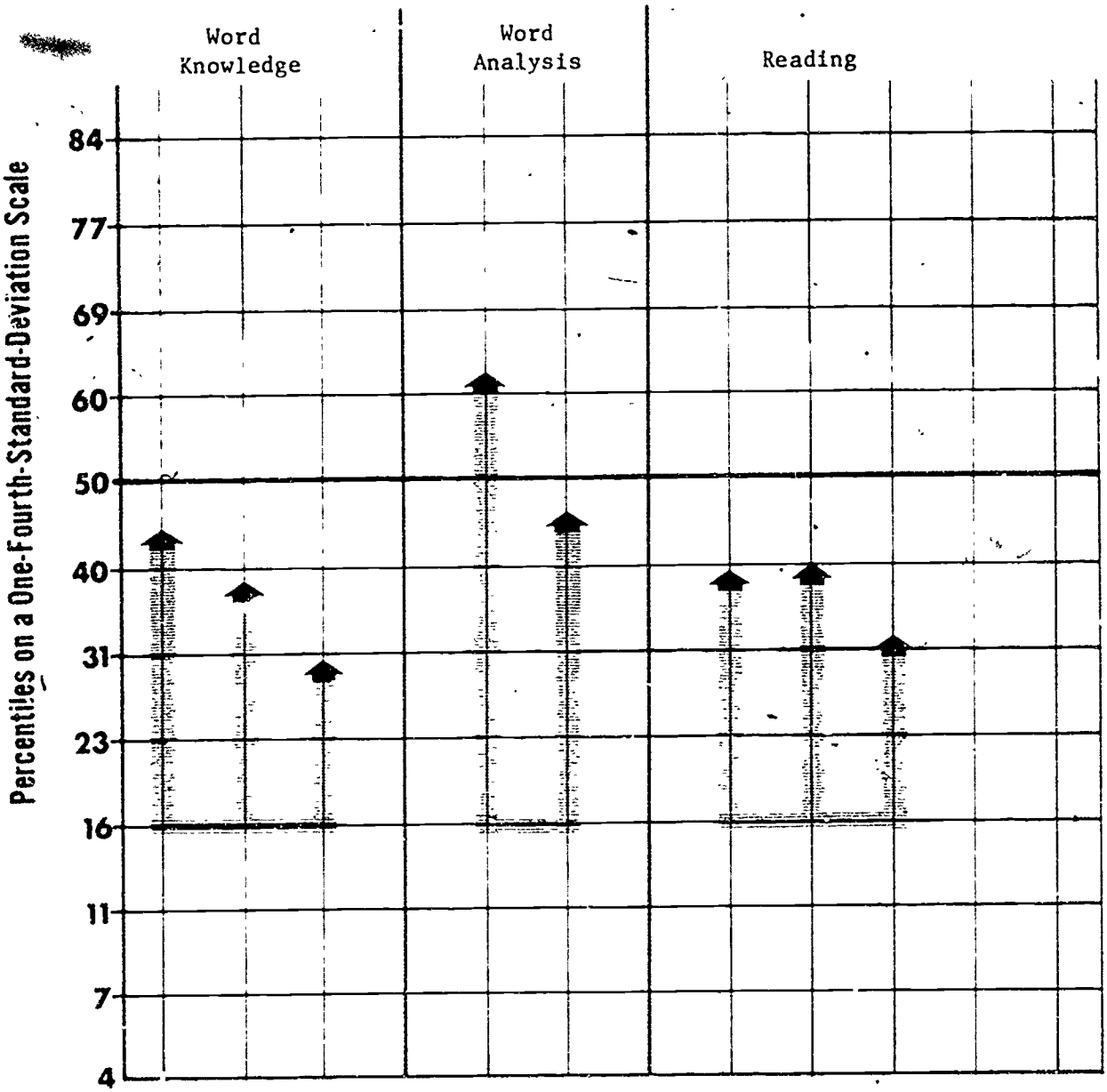
- - - - - - -

Table 3.12

- - - - - -

Table 3.10

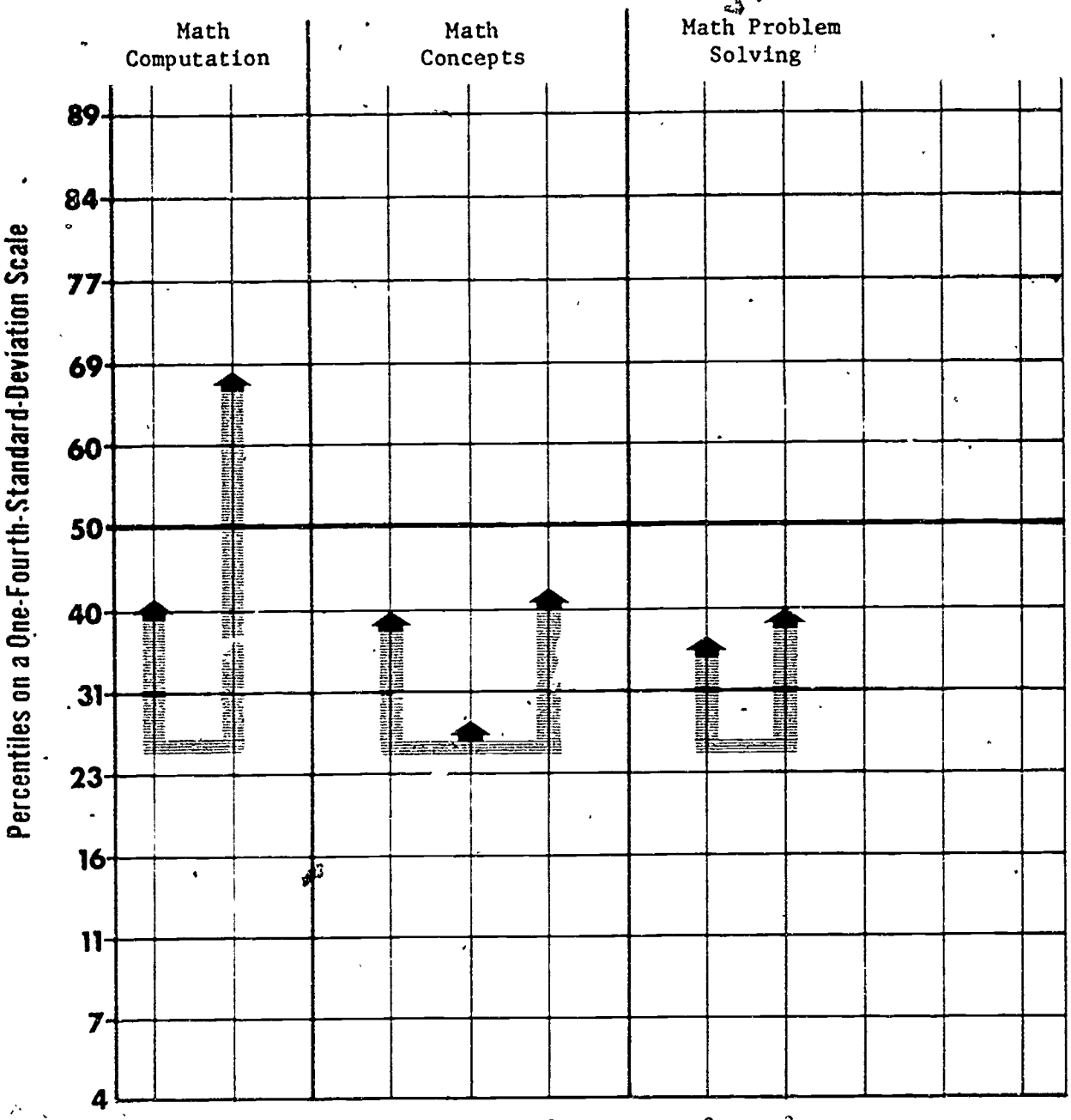
Norm-Referenced Gains on MAT Reading Subscales by Grade Level for 1st-Starting, Low-Income, E-B Follow Through Students



POST	1	2	3	1	2	1	2	3
N	1268	1687	1933	1270	1686	1271	1687	1943
%-tile	44	38	30	62	46	39	40	32
S. S.	39.0	52.2	56.9	39.7	52.0	37.4	51.1	55.6
S. D.	10.0	7.5	8.3	6.8	8.7	9.0	8.4	10.1
G. E.	1.87	2.65	3.08	1.91	2.83	1.74	2.58	3.00

Table 3.11

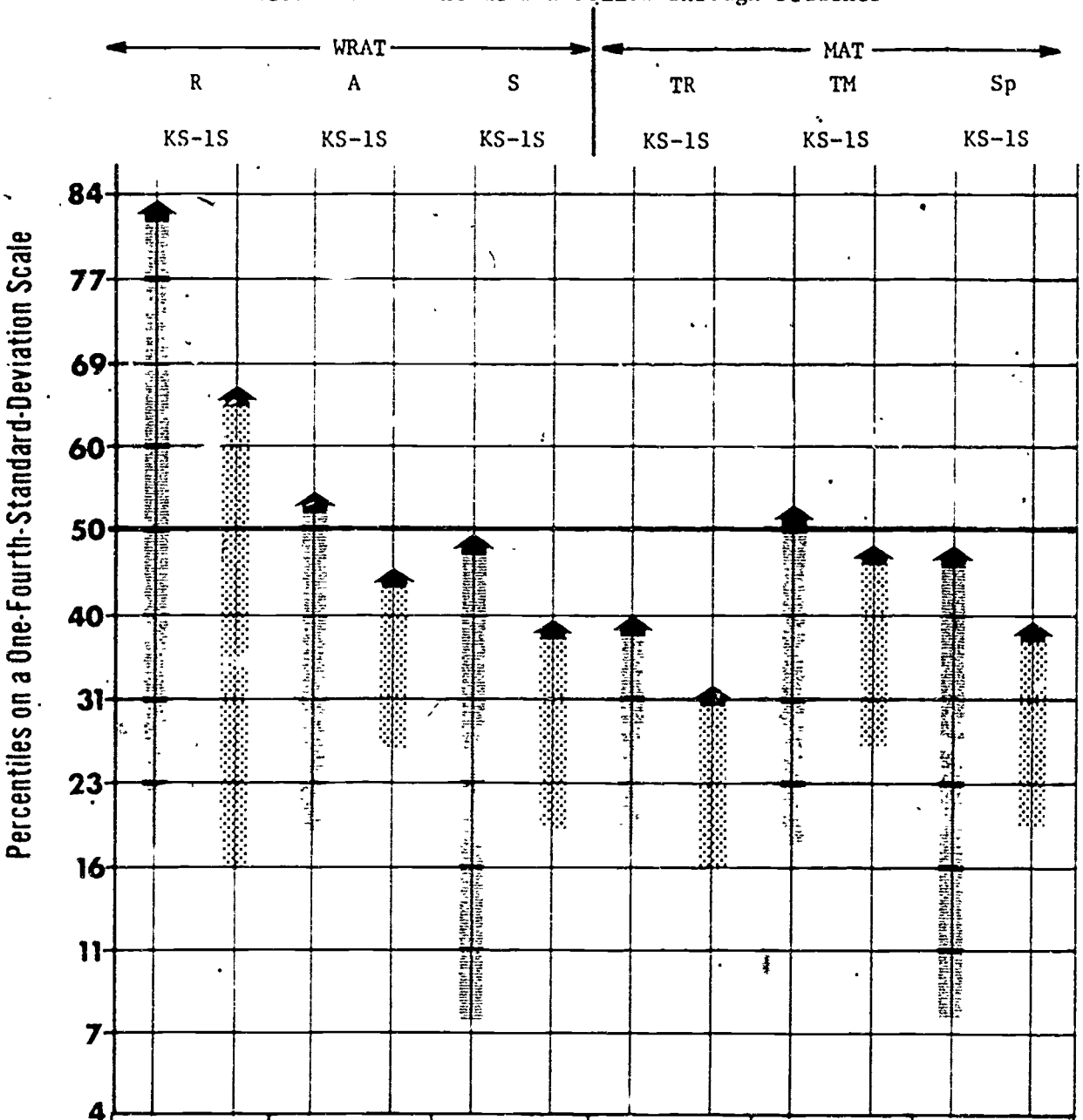
Norm-Referenced Gains on MAT Math for
1st-Starting, Low-Income, E-B Follow Through Students



POST	2	3	1	2	3	2	3
N	1683	1938	1269	1685	1934	1684	1935
%-tile	41	68	40	28	42	37	40
S. S.	52.2	69.8	38.0	51.2	64.1	53.3	65.0
S. D.	9.0	11.7	10.4	9.7	10.9	9.9	11.4
G. E.	2.73	4.30	1.68	2.57	3.71	2.68	3.64

Table 3.12

A Comparison of Norm-Referenced Gains on the WRAT and MAT from K-Starting and 1st-Starting Sites for Low-Income E-B Follow Through Students



	R		A		S		TR		TM		Sp	
PRE	K	1	K	1	K	1	K	1	K	1	K	1
POST	3	3	3	3	3	3	3	3	3	3	3	3
N	1988	3629	1989	3630	1974	3609	1863	1943	1804	1926	1824	1863
%-tile	83	67	54	45	49	39	40	32	52	48	48	39
S. S.	114.5	106.5	101.2	97.9	99.7	95.9	58.0	55.1	71.1	69.9	62.8	60.5
S. D.	20.9	19.2	8.6	9.8	13.9	15.1	10.3	10.1	11.9	11.1	11.3	10.8
G. E.	5.35	4.66	3.99	3.78	3.79	3.59	3.31	3.00	3.98	3.84	3.68	3.41

Effects of Placing Retentions with Agemates Versus Grademates

In the data presented so far, we have placed children who have been retained one grade or more with their grademates. We have also analyzed the data keeping retentions with their agemates (their starting cohort). For a norm-referenced comparison, it makes sense to keep retentions with their grademates, since norms are built with this procedure. For K-starting Low-Income students, retentions were 6.2 percent. For 1st-starting Low-Income students, retentions were 11 percent. In some of our Spanish speaking sites it was customary to retain over 60 percent of the students in first grade prior to Follow Through. We could find no national statistics on retentions. However, judging from percent of children above the modal age of fourth graders, we would estimate retentions to fall between 8 and 12 percent nationally. In general, as sponsor, we have discouraged retentions until this year. We now encourage keeping the children back until they complete three full years of the Distar programs.

Table 3.13 shows that the effect of placing retentions with grademates is to increase the total number of children in the analysis at the end of third grade, but not to appreciably change the findings. On the WRAT, there is at most the increase of one standard score unit (one-fifteenth standard deviation unit) when retentions are placed with grademates. This is a negligible effect. On the MAT the changes are variable. The mean standard score for Language goes down for K-starting students. Reading and Math means do not change for K-starting students, but show a small increase for 1st-starting students.

- - - - - -

Table 3.13

- - - - - -

Table 3.13

Effects of Placing Retentions with Agemates Versus Grademates (Low-Income students only)*

K-Starting

Variable	With Grademates			With Agemates		
Post 3rd Grade	Mean S.S.	S.D.	N	Mean S.S.	S.D.	N
WRAT						
Reading	114.0	20.5	1911	113.0	20.2	1866
Arithmetic	101.2	8.6	1912	100.9	8.6	1867
Spelling	99.3	13.3	1897	98.9	13.2	1851
MAT						
Total Reading	58.0	10.3	1784	58.0	10.5	1593
Total Math	71.0	12.0	1727	70.8	12.8	1547
Spelling	62.8	11.2	1708	62.4	11.5	1532
Language	69.9	13.2	1775	71.3	13.0	1470
WRAT		1st-Starting				
Reading	105.3	18.4	3460	104.8	18.0	3224
Arithmetic	97.5	9.4	3461	96.8	9.7	3230
Spelling	95.0	13.9	3440	94.7	13.8	3216
MAT						
Total Reading	55.0	8.9	1883	54.0	8.9	1525
Total Math	69.8	11.1	1878	67.7	12.5	1519
Spelling	60.4	10.8	1816	59.5	11.1	1466
Language	70.6	12.1	1893	70.1	12.3	1378

*These N's for students with Grademates are slightly lower than those in Tables 3.1 and 3.7 which have similar data. This is because the effects of retention placement comparisons were made prior to a final data correction which brought Head Start Planned Variation children into the summary analysis, and in one site children grouped as "unknown" were identified as Low-Income and moved into that analysis. Thus, the differences reflect a conscious effort to insure that a fair comparison is being presented.

With these data available, the reader can be assured that a distortion of the findings has not been produced by the decision on where to place retentions.

The Effects of Adding Non-Low-Income Students

One would expect the 20 percent Non-Low-Income students in our Follow Through classes would do better than their Low-Income peers. Table 3.14 shows that when Non-Low-Income students are added to the analysis, the means go up on nearly all variables one to two standard score units. Standard deviations also increase slightly. These outcomes are not surprising. An estimate of the mean standard score for Non-Low-Income students can be crudely obtained by multiplying the differences shown by five and adding the result to the mean for Low-Income students.

- - - - - -

Table 3.14

- - - - - -

A Comparison of Three Types of Analysis of Gains

As noted in Chapter 2, the "gains" records were analysed in three ways. First, the summary statistics were based on all available records in the "gains" file. This is the data base we have been looking at so far (the MAX-N GAINS). It is fair to question whether this data base, even with the large N's, truly reflects student gains within the program. To answer this question, two additional types of analyses were made. First, we analyzed BETWEEN-YEAR GAINS for all Low-Income students in the "gains" file. Second, we examined the outcome for students on whom we had both entry and exit tests (PRE-POST GAINS). For K-starting sites, this N is very small for three reasons: (1) early cohorts were not tested at entry, especially where SRI was testing, (2) two later cohorts have not yet finished the program, and (3) standard score norms on the

Table 3.14

Effects of Adding Non-Low-Income Students

(Retention with grademates)

K-Starting

Variable	Low-Income			All		
Post 3rd Grade	Mean S.S.	S.D.	N	Mean S.S.	S.D.	N
WRAT						
Reading	114.5	20.9	1988	115.5	21.1	2308
Arithmetic	101.2	8.6	1989	101.7	8.7	2311
Spelling	99.7	13.9	1974	100.3	14.0	2296
MAT						
Total Reading	58.0	10.3	1863	58.8	10.7	2127
Total Math	71.1	11.9	1804	71.8	12.0	2063
Spelling	62.8	11.3	1824	63.2	11.2	2017
Language	69.9	13.2	1854	71.0	13.6	2118
WRAT	1st-Starting					
Reading	106.5	19.2	3629	107.6	19.5	4432
Arithmetic	97.9	9.8	3630	98.6	9.9	4431
Spelling	95.9	15.1	3609	97.1	15.7	4406
MAT						
Total Reading	55.6	10.1	1943	56.2	9.8	2358
Total Math	69.9	11.1	1926	71.2	11.5	2354
Spelling	60.5	10.8	1863	61.3	11.0	2285
Language	70.7	12.2	1941	72.1	12.8	2371

WRAT are lacking for pre-K children under 5 years of age. These three factors combine to give a low N on WRAT standard scores for K-starting PRE-POST GAINS (N = 169 to 178).

Table 3.15 shows the results for the three types of analysis for K-starting students on WRAT Reading and Arithmetic. A very interesting effect is present in the BETWEEN-YEARS GAIN data. The Post-score group for a lower grade level always has a higher mean than the Pre-score group for a higher grade level. For example, post-kindergarten mean on Reading for 1,287 students was 110.5. The pre-first mean for 3,766 students was 107.4. This effect is due to the fact that there are relatively more students in lower-grade-level groups from later-starting cohorts. Later cohorts do progressively better.

- - - - - -

Table 3.15

- - - - - -

When the three kinds of analyses are compared, the following conclusion is apparent: The MAX-N GAIN analysis conservatively represents the true gains made by Low-Income Follow Through students in the E-R program. If we add together the BETWEEN-YEAR GAINS, and add this sum to the entry score, the gain would be much greater than that shown in the MAX-N GAIN analysis. The PRE-POST GAIN also shows a slightly larger gain than the MAX-N GAIN.

Since the MAT was not given at entry, and not used at all until 1972, a similar analysis cannot be made for MAT scores. However, for the 375 students in the PRE-POST GAIN analysis, the mean post-third-grade score was 59.6 on MAT Total Reading. This score is 1.6 standard

Table 3.15

Comparison of MAX-N GAIN Analysis Effects with those for BETWEEN-YEAR GAIN and PRE-POST GAIN Analyses on K-Starting Low-Income Students[1]

WRAT Reading

	MAX-N GAIN[2]			BETWEEN-YEAR GAIN[2]			PRE-POST GAIN[3]		
	Mean S.S.	S.D.	N	Mean S.S.		N	Mean S.S.	S.D.	N
Pre K	86.1	16.0	1948	86.6			85.4	13.3	169
						1287			
Post K	107.5	18.5	4621	110.5	107.4				
						3766			
Post 1	115.2	20.1	4378	116.2	114.5				
						2825			
Post 2	112.8	19.7	3023		113.1	112.4			
						1770			
Post 3	113.1	20.2	1866		113.1		113.2	20.0	169

WRAT Arithmetic

	MAX-N GAIN[2]			BETWEEN-YEAR GAIN[2]			PRE-POST GAINS[3]		
	Mean S.S.	S.D.	N	Mean S.S.		N	Mean S.S.	S.D.	N
Pre K	86.6	15.7	2141	87.0			86.1	16.0	178
						1398			
Post K	108.2	14.4	4628	110.7	108.2				
						3767			
Post 1	105.5	11.2	4378		106.2	105.4			
						2816			
Post 2	99.7	9.3	3014		99.8	99.3			
						1775			
Post 3	100.9	8.6	1867		100.9		100.7	7.4	178

[1]As with Table 3.13, this analysis was completed prior to making a few final adjustments to included Planned Variation Head Start kids and to correct the placement of a group of Low-Income kids whose income level had not previously been identified correctly.

[2]For these data, retentions are grouped with agemates.

[3]For these data, retentions are grouped with grademates.

score points higher than the mean for the comparable MAX-N GAIN group. For MAT Total Math, the comparison is 74.1 for the PRE-POST GAIN group and 71.0 for the comparable MAX-N GAIN group. These findings again support the conservative nature of the MAX-N GAIN report.

Table 3.16 presents a similar comparison for first-starting sites. All of the comparisons shown, including those for the MAT again show that the MAX-N GAINS present a conservative picture of the true gains.

- - - - - -

Table 3.16

- - - - - -

Note: Since we do not have available the BETWEEN-YEARS GAIN with retentions placed with grademates, the data for the MAX-N Analysis in Tables 3.15 and 3.16 is shown for retentions with agemates. As should be apparent from Table 3.13, the effects of these different placements of retentions are minor and do not alter the conclusion drawn above.

Comparison of Findings for "Singles" and "Later Entering" Groups

As noted in Chapter 2, we have about 4800 student records where there is only testing at one point in time ("Singles") and thus no basis for computing gains. These records arise for several reasons. In the early years, many students were not pretested, were given a post

Table 3.16

Comparison of MAX-N GAIN Analysis Effects with those for BETWEEN-YEAR GAIN and PRE-POST GAIN Analyses on 1-at Starting Low-Income Students[1]

WRAT Reading

	MAX-N GAIN[2]			BETWEEN-YEAR GAIN[2]		PRE-POST GAIN[3]		
	Mean S.S.	S.D.	N	Mean S.S.	N	Mean S.S.	S.D.	N
Pre 1	82.1	11.4	1746	82.4		82.2	11.4	837
Post 1	98.6	17.1	4111	100.9 99.2	1549			
Post 2	103.7	18.1	4007	104.4 104.2	3522			
Post 3	104.8	18.0	3224	105.2	2987	106.3	16.1	837

MAT Total Reading

Post 3	55.0[3]	8.9	1883			56.5[3]	8.1	818

WRAT Arithmetic

	MAX-N GAIN[2]			BETWEEN-YEAR GAIN[2]		PRE-POST GAIN[3]		
	Mean S.S.	S.D.	N	Mean S.S.	N	Mean S.S.	S.D.	N
Pre 1	88.3	12.3	1775	88.7		89.8	12.0	857
Post 1	97.7	11.8	4087	100.0 98.0	1568			
Post 2	94.3	8.9	4005	94.7 94.2	3496			
Post 3	96.8	9.7	3230	97.1	2989	100.4	8.6	857

MAT Total Math

Post 3	69.8[3]	11.1	1878			72.7[3]	10.0	816

[1]Data based on slightly smaller sample size than in Table 3.7. See footnote to Table 3.13.

[2]For these data, retentions are grouped with agemates.

[3]For these data, retentions are grouped with grademates.

kindergarten test and then left the program to attend parochial school. In Grand Rapids two whole cohorts were pretested in the fall and not tested in the spring because of a lack of a functioning memorandum of agreement. Many students moved from the program or were migrants. Finally, bad testing procedures in one site early in the program led to a large number of single records after some tests were discarded.

The present data on students where there are single records does not differ appreciably from that reported in Technical Report 73-2, December 1973. The students tend to show lower test performances in proportion to their lower placements in the Distar programs. The same conclusion can also be drawn for the 1900 "later entering" students in our file. The students perform more poorly when they have not been full-term in the E-B Follow Through program. This is to be expected if the program is important to their success.

IMPROVEMENT IN IQ

The Slosson Intelligence Test was used as a measure of general language competency and more generalized cognitive skills. The findings generally show an improvement in IQ from entry to end of third grade. In K-sites (Table 3.17), the average improvement is 2.3 points for low-income students in the PRE-POST GAIN analysis. The BETWEEN-YEARS GAIN analysis shows a gain of 9.6 points during the kindergarten year and small losses totaling 3.6 points after that. The MAX-N GAIN analysis shows a gain of 4.5 points. The PRE-POST GAIN analysis utilizes the data from cohorts 3 and 4 exclusively. It appears that cohorts 5 and 6 are showing gains that will be closer to 6 or 7 points than 2.

- - - - - -

Table 3.17

- - - - - -

Table 3.17

Slosson IQ Gains, K-Starting Students

PRE-POST GAIN

	Retentions with Grademates				
	Pre K	Post 3	Gain	S.D.	N
All	106.6	109.2	2.55*	15.5	346
Low Income	105.7	108.0	2.30*	15.5	293

*Significant beyond .01 level.

BETWEEN-YEARS GAINS

	Retentions with Agemates				
Low Income Only	Pre	Post	Gain	S.D.	N
Pre K - Post K	103.2	112.6	9.40**	14.2	1850
Post K - Post 1	110.2	108.8	-1.45**	13.1	2495
Post 1 - Post 2	109.8	109.4	- .04	12.9	1079
Post 2 - Post 3	110.3	108.5	-1.81**	11.6	791

**Significant beyond .001 level.

MAX-N GAIN

	Retention with Grademates		
Low Income Only	Mean	S.D.	N
Pre K	102.5	14.6	2642
Post K	109.8	15.2	4509
Post 1	108.1	14.8	3174
Post 2	09.1	16.6	2068
Post 3	106.9	16.7	1933

* $p < .01$

** $p < .001$

When examined more closely, it is apparent that the IQ gains for K-sites are quite variable. With 73 students in the PRE-POST GAINS analysis, Cherokee shows a mean gain of 9.4 points (101.1 to 110.5). On the other hand, West Iron County shows a loss of 9.8 points (111.8 to 102.0) for 19 students in cohort 3. This group was pulled out of the program before they finished it. In general, gains and losses seem related to adequacy of program implementation.

The IQ gains for 1st-sites are more substantial. The average is 5 to 6 points no matter which gains analysis is examined. This gain is over one-third standard deviation. Examination of the data by site suggests the importance of program implementation. For example, in Las Vegas, New Mexico, after two years of modest gains (4.1 and 3.8), a push on taking the kids as far as they could go, produced a mean gain of 13 points for 66 children. Larger gains also tend to be found in areas where the parent language is not English (Rosebud, Uvalde, Dimmitt).

- - - - - -

Table 3.19

- - - - -

The effects of the gains in 1st-starting sites is to bring the mean student IQ to the average for the general population.

Distributional Analyses

It is possible for mean effects to be distorted by unusual distributions. This is not the case for the present data. When we examine the distribution of students by scatines and the percent of students above grade norm, the results are very consistent with those already presented. Table 3.20 presents the per cent of children above grade norm

Table 3.18

Slosson IQ Gains, 1st-Starting Students

PRE-POST GAINS

	Retentions with Grademates				
	Pre K	Post 3	Gain	S.D.	N
All	95.5	101.9	6.38*	14.5	1208
Low Income	94.7	101.0	6.31*	14.6	996

*Significant beyond .001 level.

BETWEEN-YEAR GAINS

	Retentions with Agemates				
	Pre	Post	Gain	S.D.	N
Low-Income Student					
Pre 1 - Post 1	93.6	98.6	4.96*	11.5	1592
Post 1 - Post 2	97.7	99.3	1.61*	12.1	2533
Post 2 - Post 3	98.8	99.3	.48*	11.9	2154

*Significant beyond .001 level.

MAX-N GAIN

	Retention with Grademates		
Lo. Income Student	Mean	S.D.	N
Pre 1	94.1	14.5	2639
Post 1	97.1	14.7	4560
Post 2	98.6	15.4	3538
Post 3	99.4	16.8	3651

for the students in the MAX-N GAIN analysis (retentions with grademates). The analysis shows that more than 50 percent of our K-starting Low-Income children are above grade norms on each measure except WRAT Spelling and MAT Reading. The 1st-starting students show a lower percentage above grade norms except on MAT Math and Language where they do as well. The importance of using the kindergarten year for instruction is as apparent in these data as in the data based on means.

- - - - - -

Table 3.19

- - - - - -

Figures 3.1, 3.2, and 3.3 show the distributions by stanine groupings for ALL students in the MAX-N GAIN analysis for the Pre-K and Post-3rd testings. The distributions for WRAT Reading show a remarkable shift to higher stanines. At pre-kindergarten 58 percent of the children fell in stanines 1 to 3. Post-third grade, only 8 percent fall in stanines 1 to 3. At pre-kindergarten, only 6 percent of the children fell in stanines 7 to 9. Post-third grade, 53 percent fall in these top level stanines. Strong distributional shifts are also present for Arithmetic and Spelling, but they are not as pronounced as that for Reading.

- - - - - - - - - - -

Figures 3.1, 3.2, 3.3

- - - - - - - - - - -

The percentage of cases by stanines on the MAT tests for K-starting children at the end of first, second, and third grades are given in Tables 3.20, 3.21, and 3.22. Our Follow Through children do better than the norm group on Total Reading at the end of first and second grades, but fall somewhat behind on the third grade test. This is

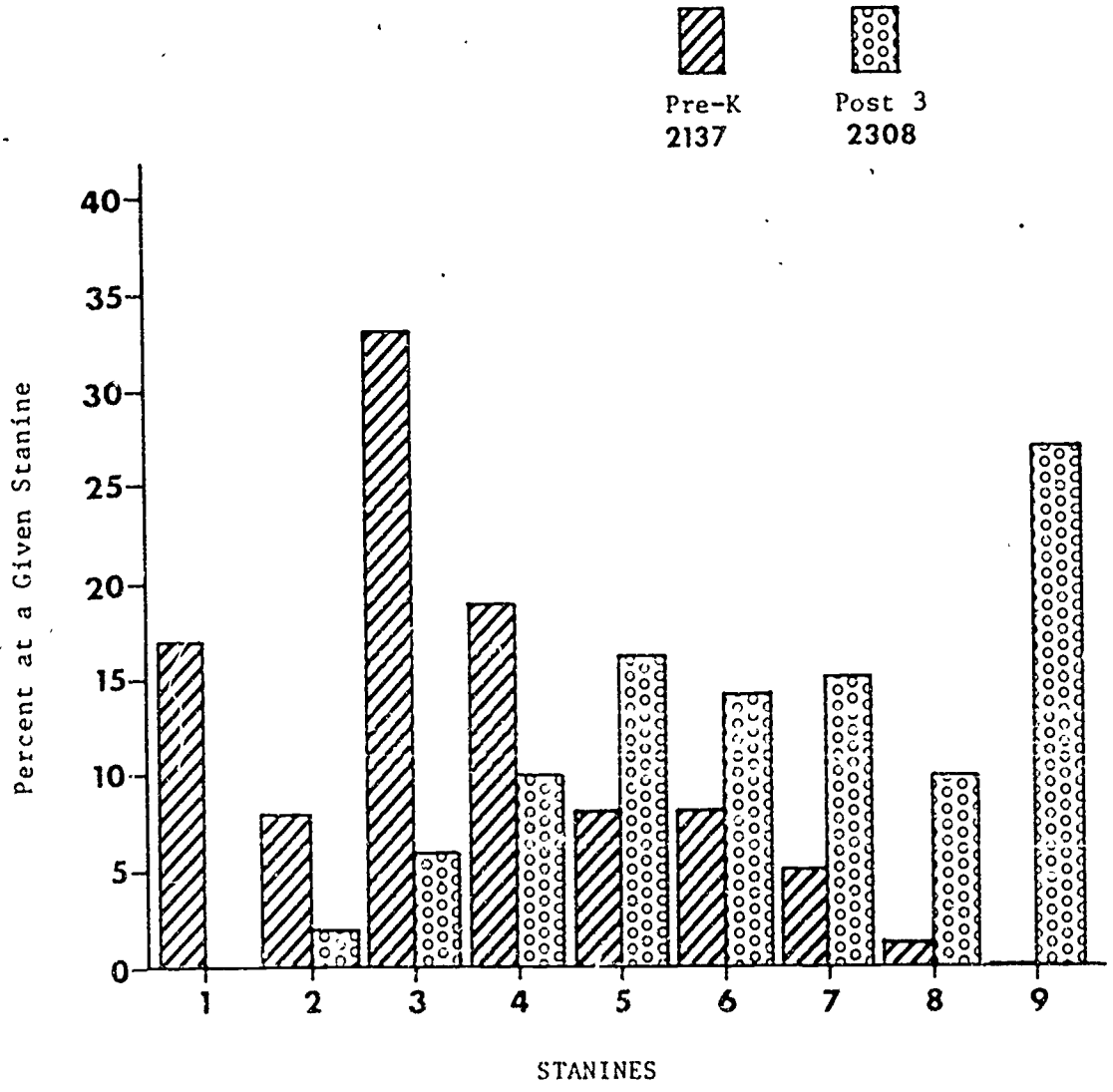


Figure 3.1 Distribution on WPAT Reading by stanines for Follow Through Students (All) in MAX-N GAIN analysis for Pre-K and Post 3rd testing.

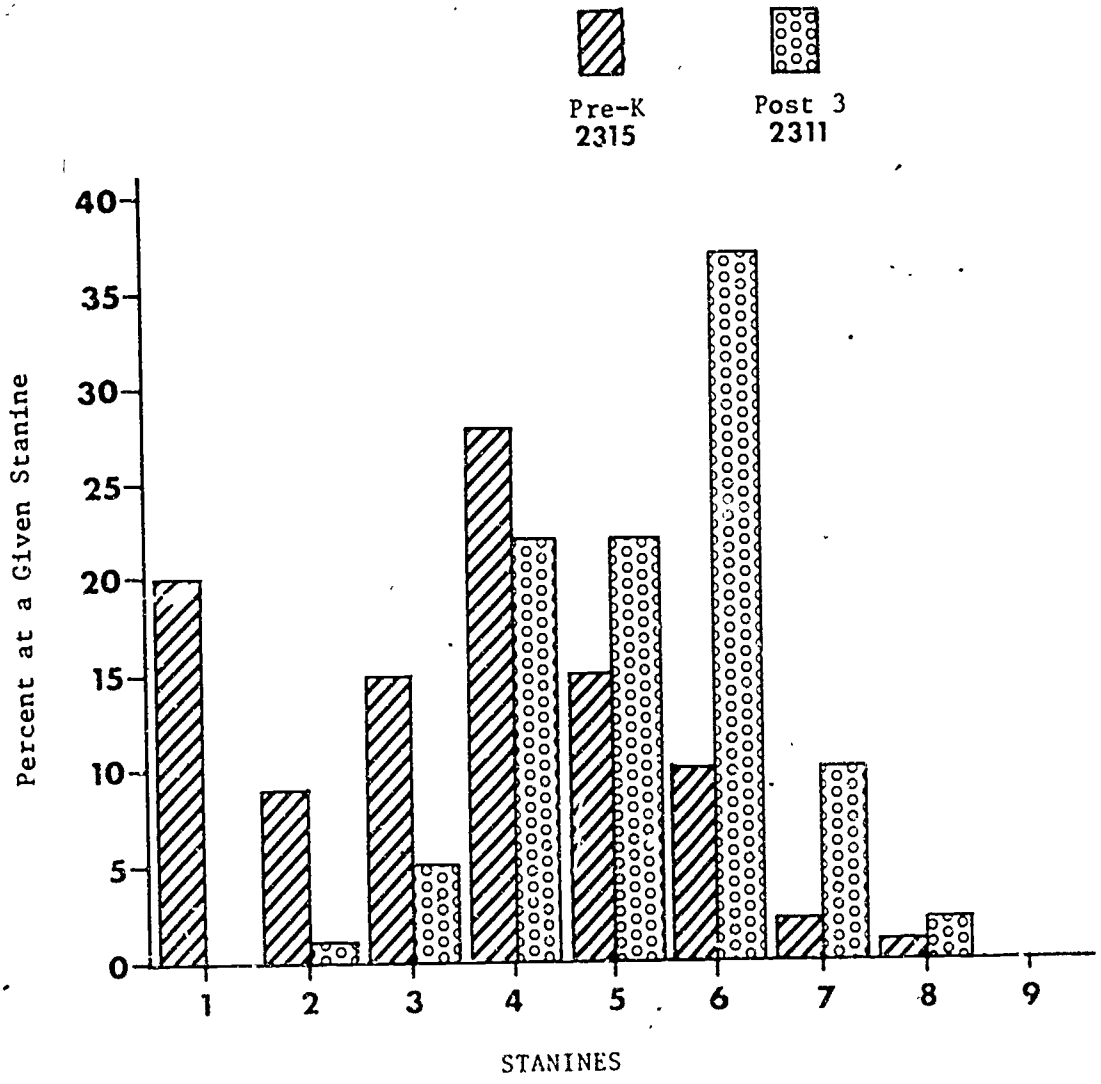


Figure 3.2 Distribution on WRAT Arithmetic by stanines for Follow Through Students (All) in MAX-N GAIN analysis for Pre-K and Post 3rd testing.

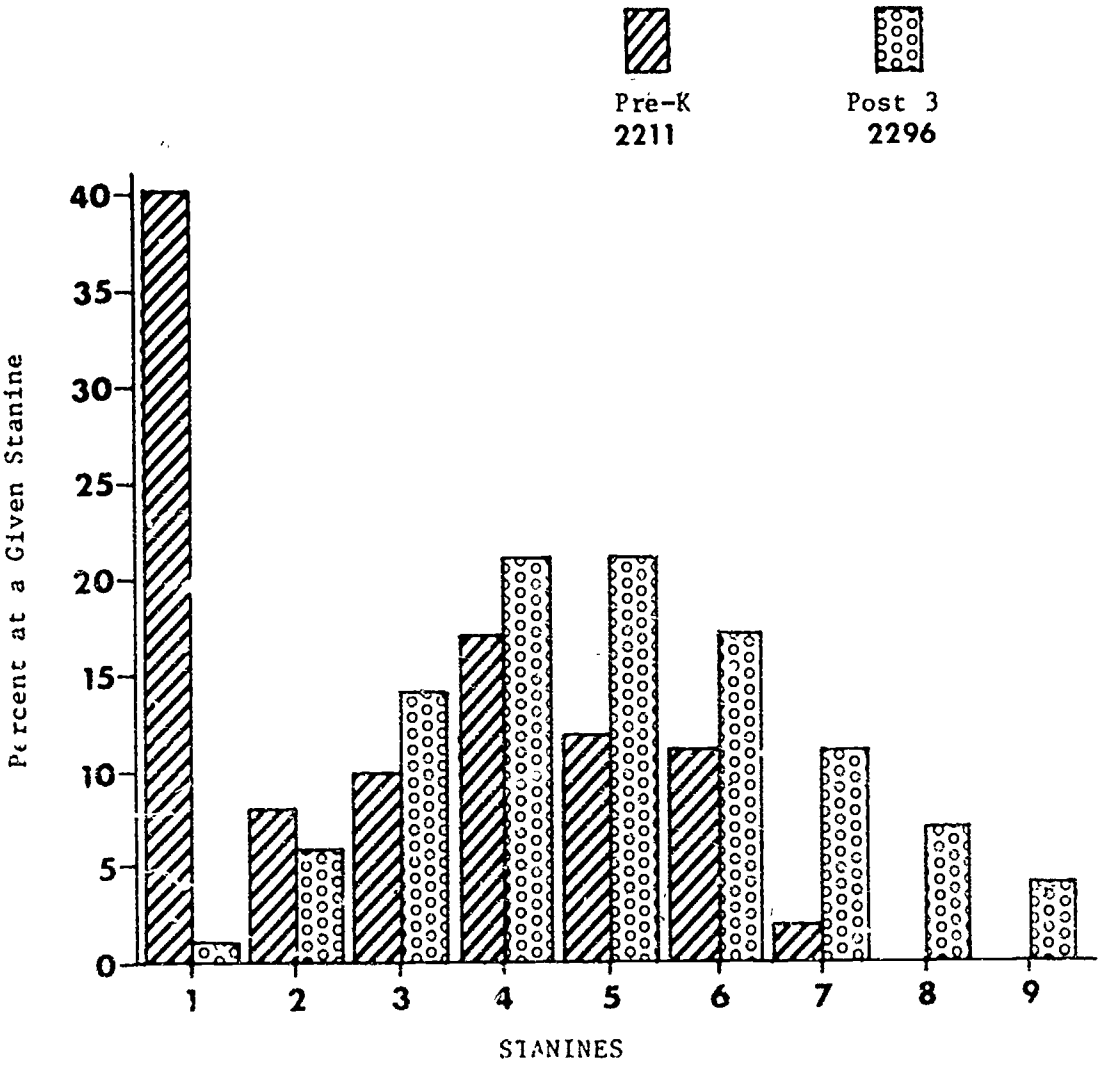


Figure 3.3 Distribution on WRAT Spelling by stanines for Follow Through Students (All) in MAX-N GAIN analysis for Pre-K and Post 3rd testing.

Table 3.19

Percent of Children above Grade Norm

Variable	Low Income N Above	Low Income N Total	Low Income %	All N Above	All N Total	All %
K-Starting						
WRAT						
Reading	1439	1890	76.1	1774	2308	76.9
Arithmetic	1177	1893	62.2	1452	2311	62.8
Spelling	854	1879	45.4	1052	2296	45.8
MAT						
Total Reading	716	1777	40.3	882	2127	41.5
Total Math	925	1719	55.0	1138	2063	55.2
Spelling	866	1697	51.0	1046	2017	51.9
Language	933	1770	52.7	1156	2118	54.6
1st Starting						
WRAT						
Reading	2099	3499	60.0	2770	4432	62.5
Arithmetic	1515	3500	43.3	2025	4428	45.7
Spelling	1090	3481	31.3	1505	4406	34.2
MAT						
Total Reading	480	1903	25.2	674	2356	28.6
Total Math	971	1904	51.0	1283	2353	54.5
Spelling	743	1841	40.4	985	2283	43.1
Language	1064	1916	55.5	1385	2371	58.4

consistent with findings reported earlier. The new information to note is that in the low stanines (1, 2, 3) we match the norm group quite well. It is only in the high stanines (7, 8, 9) that we fall behind in frequency of cases. The implication is that the program produces students with good reading comprehension, but we lack our share of the top level kids in language competency. This is to be expected if disadvantaged children are on the average weak in language competencies. Our program compensates for this by building a basic understanding of the language of instruction, but cannot in the time available build adult-like vocabularies for some of these children. The latter apparently occurs in many middle-class homes.

On the MAT Total Math test (Table 3.21), percentage of cases by stanines shows an improvement from first to third grade for K-starting children. Table 3.22 shows a distribution that is a little better than the norm group for Language (grammar) and a little worse for Spelling.

Table 3.20

All Children from K-Sites

MAT Total Reading Percentages at a Given Stanine

Stanine	Percent at Each Stanine			
	Post 1st	Post 2nd	Post 3rd	Norm Expectation
1	2	1	3	4
2	3 } 11	4 } 12	8 } 25	7 } 23
3	6	7	14	12
4	9	17	20	17
5	14 } 41	25 } 66	26 } 62	20 } 54
6	18	24	16	17
7	22	14	7	12
8	14 } 48	4 } 22	4 } 13	7 } 23
9	12	4	2	4
	N = 3183	2556	2127	

Table 3.21

All Children from K-Sites

MAT Total Math Percentages at a Given Stanine

Stanine	Percent at Each Stanine: Post 1st		Post 2nd		Post 3rd		Norm Expectation	
1	2		9		4		4	
2	9	24	9	26	7	20	7	23
3	13		14		9		12	
4	21		21		16		17	
5	22	59	17	52	17	51	20	54
6	16		14		18		17	
7	8		13		17		12	
8	6	17	5	23	8	30	7	23
9	3		5		5		4	

N = 3207

Table 3.22

All Children from K-Sites

MAT Language and Spelling Percentages at a Given Stanine

Stanine	MAT Language: Post 3rd		MAT Spelling: Post 2nd		Post 3rd		Norm Expectation	
1	4		1		3		4	
2	6	20	7	20	7	22	7	23
3	10		12		12		12	
4	17		18		16		17	
5	20	56	17	57	21	58	20	50
6	19		22		21		17	
7	11		10		11		12	
8	8	24	0	23	6	20	7	23
9	5		13		3		4	

N = 2118 N = 2539 N = 2017

Summary

The data presented in this Chapter show that the objective of getting disadvantaged children up to national norm averages on basic academic skills was largely met, especially for K-starting sites. Improvement on the Slosson IQ tests was also found, especially with the initially lower 1st-starting sites.

CHAPTER 4

ANALYSES BY COHORT

The data by cohort permit an examination of the reliability of the findings over time and an examination of possible program improvements over time. As was shown in the last chapter, the MAX-N GAINS data base leads to the same conclusions as the other more refined gains analyses (with a slight reduction in magnitude of effects), while providing for a much larger sample size. For these reasons we will restrict analyses in this chapter to that data base, except for IQ data. We will further focus the analyses only on Low-Income students with retentions placed with grademates.

READING

K-Starting Sites

Yearly gains (or losses) against the norm group are shown for WRAT Reading in Table 4.1 and 4.2. In these tables, the ending level from the grade before is used as the base for change for the next grade level. The data for kindergarten show little variation in the entry performance of the children, a strong gain against the norms for all cohorts, and a three-fourths standard deviation improvement from cohort 2 to cohort 6. This improvement can be attributed largely to the improvement in teaching at the kindergarten level. In the early years, kindergarten teachers were not prepared to believe that their children could be taught to read, so they did not always "follow through" with sponsor suggestions. After a number of demonstrations of what could

be done, and a real targeting by the sponsor on kindergarten-level performance of teachers and aides, improvements were made in the number of lessons taught and the quality of the teaching.

- - - - - - - - -

Table 4.1 and 4.2

- - - - - - - - -

The data for cohort 1 stands apart from the rest. These data are from only four sites and the N is small. However, as we shall see in Tables 4.3 and 4.4, a similar effect is found for 1st-starting sites. One explanation for the cohort 1 effect is that the reading (and arithmetic) program the first year moved faster than the published version the following year which gave more attention to lower performing children.

The data over cohorts for first grade, also show rather consistent gains against the norms and that improvements over cohorts tend to be maintained. The data in Table 4.2 for second and third grades, show that there are no further gains against the norm group. On the contrary, except for cohort 1, there are slight losses in second grade and essentially no changes in third grade. The pattern for level of performance over cohorts at the end of third grade is similar to that found at the end of kindergarten.

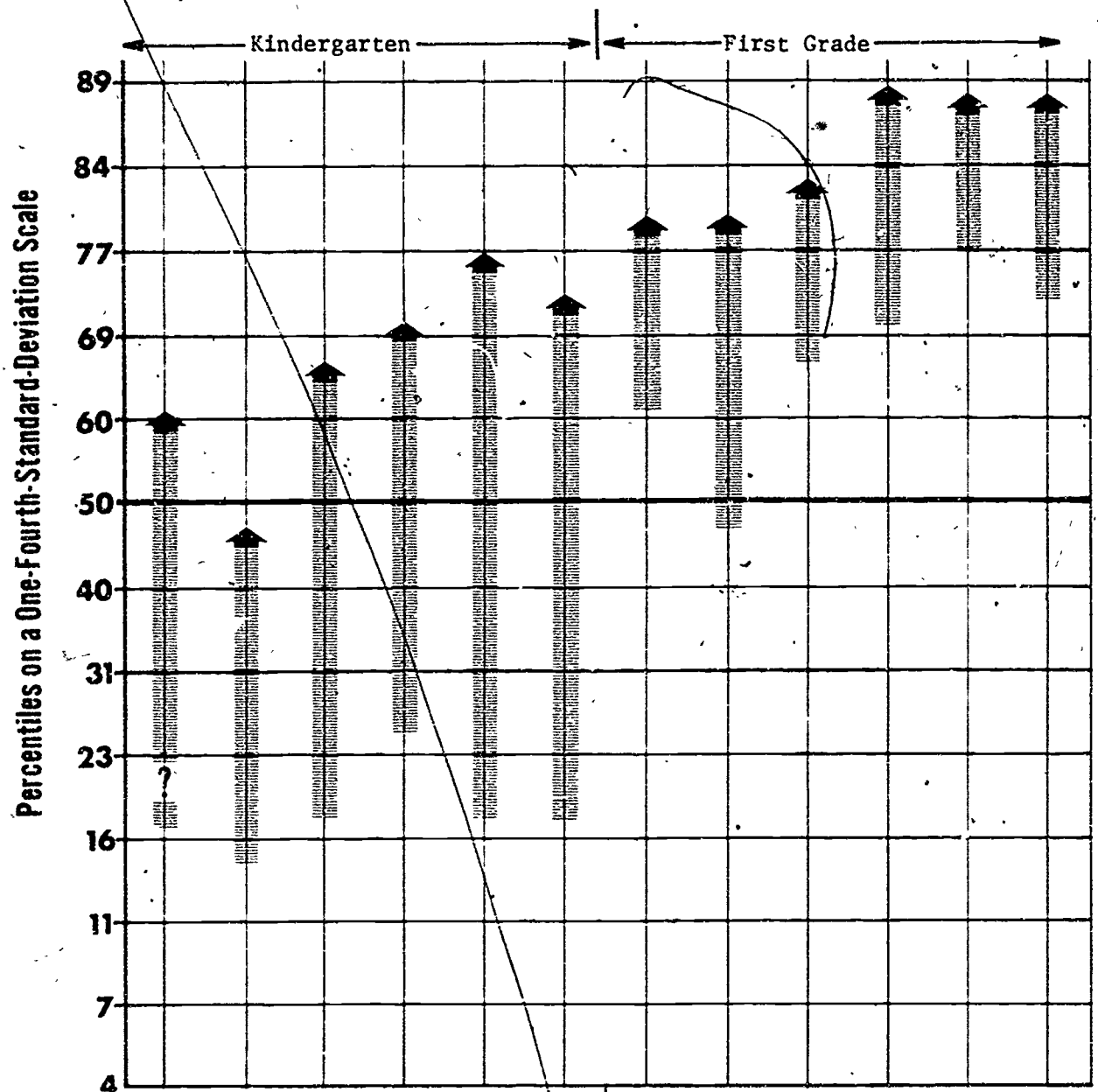
Discounting the cohort 1 data, we can conclude that the improved teaching of reading over cohorts in kindergarten has measurable effects which are still apparent at the end of third grade.

1st-Starting Sites

Tables 4.3 and 4.4 present the WRAT Reading data over six cohorts for 1st-starting sites. The tables show consistent gains against the

Table 4.1

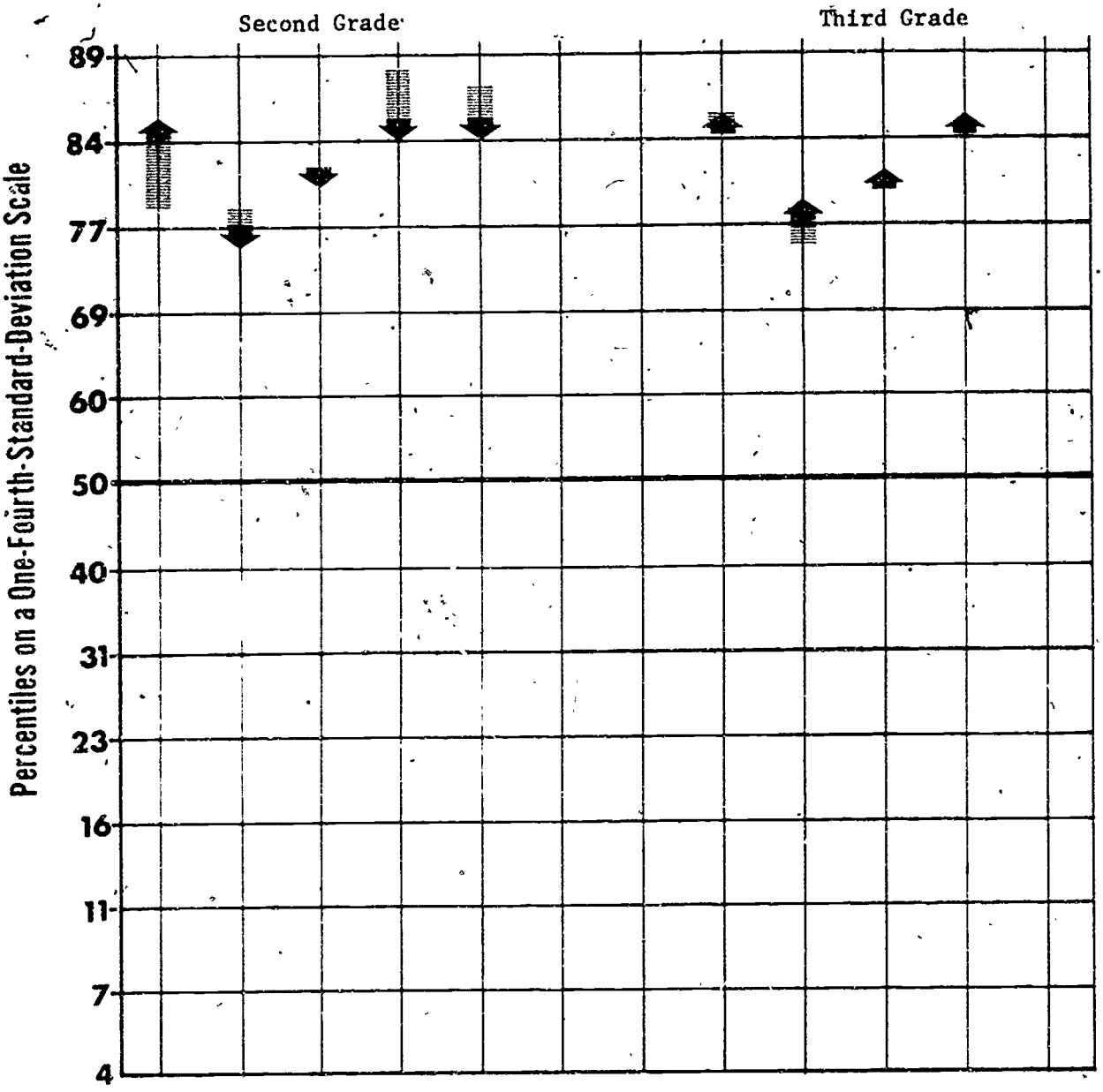
Norm-Referenced Gains on WRAT Reading by Cohort from Kindergarten through First Grade for K-Starting, Low-Income, E-B Follow Through Students



Cohort	1	2	3	4	5	6	1	2	3	4	5	6
PRE	K	K	K	K	K	K	–	–	–	–	–	–
POST	K	K	K	K	K	K	K-1	K-1	K-1	K-1	K-1	K-1
Post N	97	755	1184	1173	1045	927	281	758	1135	988	889	759
Post %-tile	61	47	66	70	77	73	79	79	82	88	87	87
Post S. S.	104	99	106	108	111	109	112	112	114	118	117	117
Post S. D.	18.9	17.9	19.6	18.5	19.0	18.3	21.1	20.6	19.5	21.2	19.9	20.4
Post G. E.	1.24	1.04	1.32	1.39	1.49	1.41	2.74	2.73	2.86	3.08	3.05	3.00

Table 4.2

Norm-Referenced Gains (Losses) on WRAT Reading by Cohort from Post-First Grade to Post-Third Grade for K-Starting, Low-Income, E-B Follow Through Students



	Cohort	1	2	3	4	5	1	2	3	4
	POST	1-2	1-2	1-2	1-2	1-2	2-3	2-3	2-3	2-3
Post	N	234	655	797	854	758	197	409	664	718
Post	%-tile	86	75	81	84	84	86	79	82	86
Post	S. S.	116	110	113	115	115	116	112	114	116
Post	S. D.	18.5	18.3	19.6	21.0	20.6	18.4	18.3	20.9	22.8
Post	G. E.	4.44	3.89	4.22	4.26	4.36	5.35	5.06	5.35	5.51

norms for first and second grades, which are maintained in third grade. The gain during the first year of instruction is particularly dramatic. However, there is no dramatic trend for an improvement over cohorts as

- - - - - - - - - - -

Tables 4.3 and 4.4

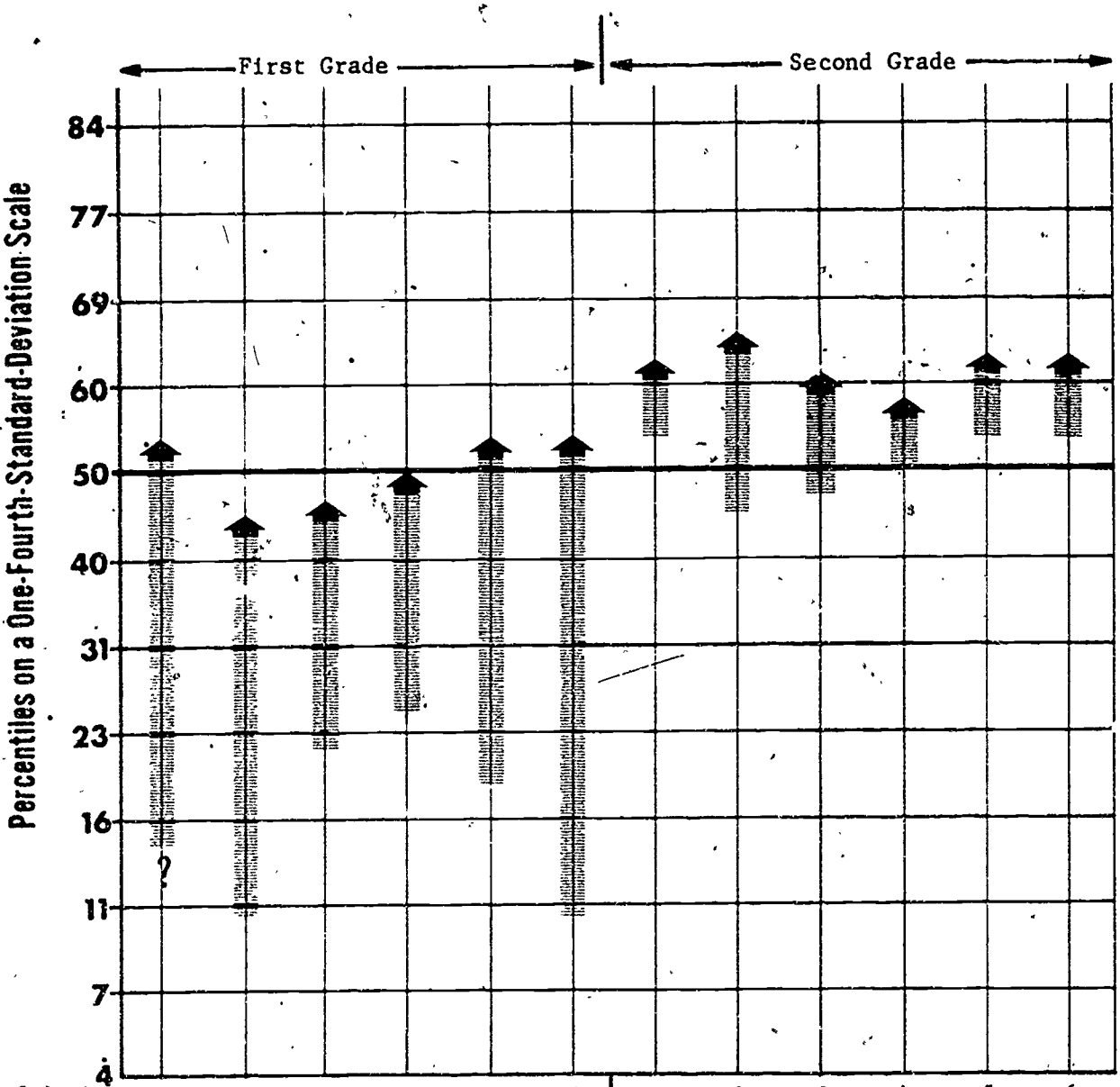
- - - - - - - - - - -

found for K-starting sites. In general, we were working with experienced teachers who were very hard workers and showed good implementation of the reading program from the start.

The high performance of cohort 1 at the end of third grade is a little perplexing. This cohort also tends to do better in spelling, and math. No MAT data are available for this cohort since they graduated before we began MAT testings. A careful examination of the data by site shows the possible contributions of several factors. First, we know in E. St. Louis that the first group was taught by the best teachers in the district and the children were strongly taught all the way through. They ended up with a mean WRAT Reading standard score of 129 ($\underline{N}$ = 166) for cohort 1. The mean for cohort 2 was "only" 115 ($\underline{N}$ = 122). Second, we note that in five other sites there are similar strong gains against the norm group from post-second to post-third grades for cohort 1 children, while the trend is for a loss against the norms the following year for cohort 2 children. We suspect that the initial excitement of a new program, combined with the field-test version of the reading program that moved more quickly than the revision must account for part of this unusual performance. Third, when the site changes are considered, we note that a high performing site that started in K and 1st simultaneously is present in cohort 1, but not

Table 4.3

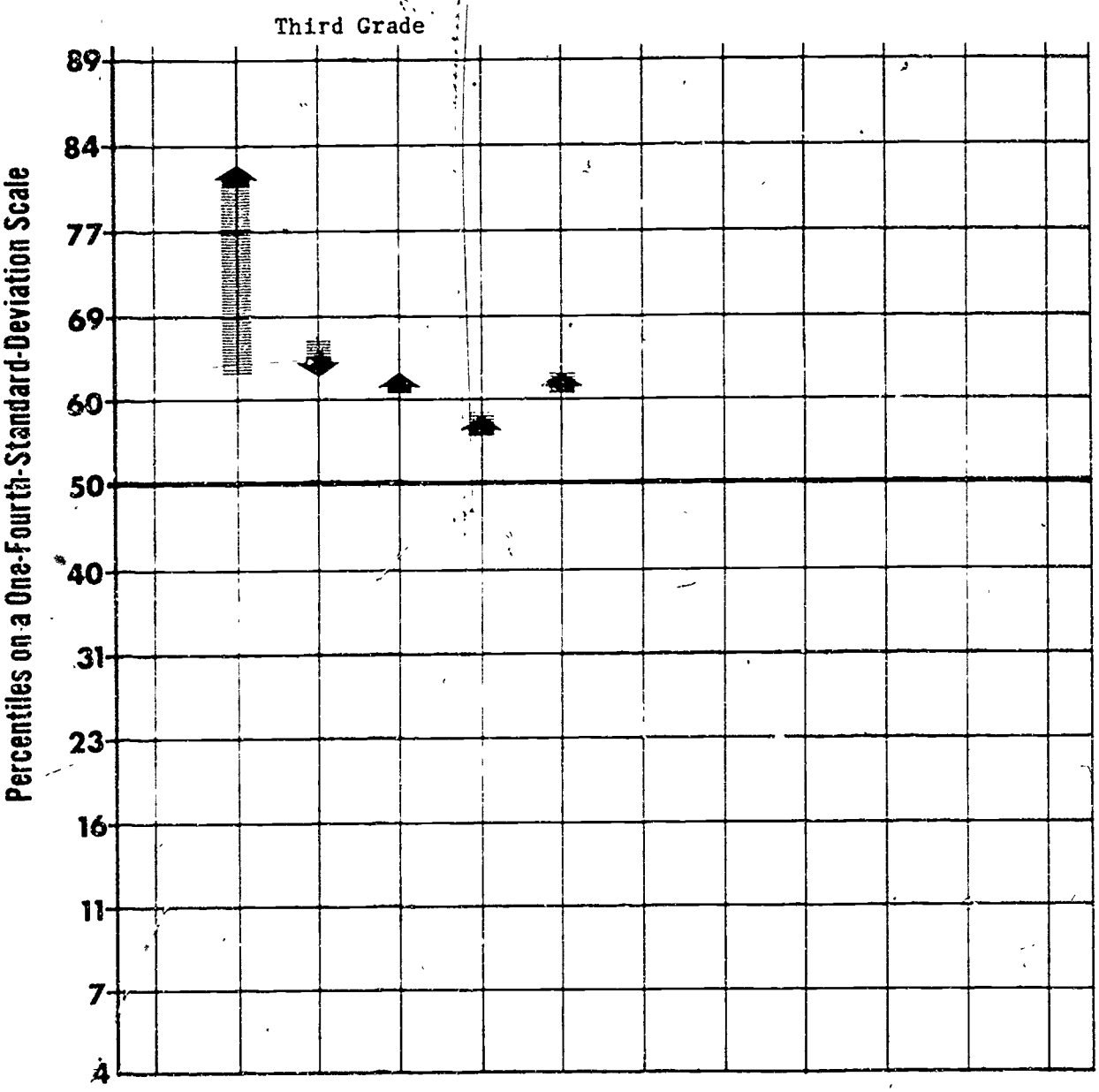
Norm-Referenced Gains on WRAT Reading by Cohort from Pre-First Grade through Second Grade for 1st-Starting, Low-Income, E-B Follow Through Students



Cohort	1	2	3	4	5	6	1	2	3	4	5	6
PRE	1	1	1	1	1	1	-	-	-	-	-	-
POST	1	1	1	1	1	1	1-2	1-2	1-2	1-2	1-2	1-2
Post N	721	1208	1009	746	713	763	775	1163	977	604	608	538
Post %-tile	53	45	47	50	53	53	63	66	61	58	63	63
Post S. S.	101	98	99	100	101	101	105	106	104	103	105	105
Post S. D.	20.0	17.7	17.9	18.9	16.6	18.1	18.7	19.1	18.4	18.0	17.0	18.2
1st G. E.	2.11	1.89	2.01	2.01	2.07	2.11	3.56	3.68	3.41	3.45	3.59	3.53

Table 4.4

Norm-Referenced Gains (or Losses) on WRAT Reading by Cohort from Post-Second Grade to Post-Third Grade for 1st-Starting, Low-Income, E-B Follow Through Students



Cohort	1	2	3	4	5
POST	2-3	2-3	2-3	2-3	2-3
Post N	678	1012	853	602	481
Post %-tile	82	63	63	58	63
Post S. S.	114	105	105	103	105
Post S. D.	22.5	18.0	19.6	18.6	16.3
t G. E.	5.48	4.50	4.50	4.43	4.44

cohort 2; while a low-performing site was new to Follow Through in Cohort 2. These three effects combine to produce an effect that is one-half standard deviation in magnitude. Note that most of the effect is accounted for by program and teacher variables. It is perhaps also important to note that cohort 1 showed significantly higher IQ gains on the SIT from the end of first grade to the end of third grade than the average of the other cohorts (96.9 to 102.0 versus 97.1 to 99.0 with N's of approximately 600 and 3000 for the two groups). (A similar comparison from the start of first grade is not possible because cohort 1 was not pretested.)

MAT

The analysis over cohorts for MAT Reading scores is not very productive, since the MAT was not added to the testing until 1972. We can note, however, that for K-starting sites in cohorts 2 and 3, there is an increase on MAT Total Reading standard scores at the end of third grade from 56.5 (N = 487) to 58.8 (N = 656), and then a slight drop for cohort 4 to 58.3 (N = 720). The lesser gain from 56.5 to 58.3 is significant beyond the .001 level (CR = 3.0) and amounts to a percentile gain from the 35th to the 41st.

For the first starting sites in cohorts 3, 4, and 5 there is a similar improvement from 54.5 (N = 854) to 54.9 (N = 593) to 55.9 (N = 482). The gain from 54.5 to 55.9 is statistically significant at the .002 level.

ARITHMETIC

K-Starting Sites

The analysis of WRAT Arithmetic data by cohorts (Tables 4.5 and 4.6) shows a very substantial improvement over cohorts in kindergarten.

Discounting the first cohort, which moved through a faster program and for which the N is small and restricted to four sites, the magnitude of the improvement is about one-half standard deviation. Although the data for the last two cohorts is not yet in, it would appear that the same magnitude of improvement over cohorts will hold at the end of third grade.

The analysis also shows a consistency across cohorts in the starting baseline performance and in the gains against the norms by grade level. On the WRAT Arithmetic tests, there is a substantial gain during kindergarten (one-and-one-half standard deviation units). A little of this gain is lost in first grade, and quite a bit is lost in second grade. But there is a recovery of small gains against the norm group in third grade, so that the net gain (K-3) against the norm group is approximately one standard deviation for later-starting cohorts. A loss in second grade (relative to the norm group) on WRAT Arithmetic has been noted in prior reports on E-B Follow Through kids. This loss is largely due to the WRAT emphasis on column functions at the second grade level, while the DISTAR program (first edition) emphasizes row functions. This difference is corrected by the end of third grade.

- - - - - - - - -

Table 4.5 and 4.6

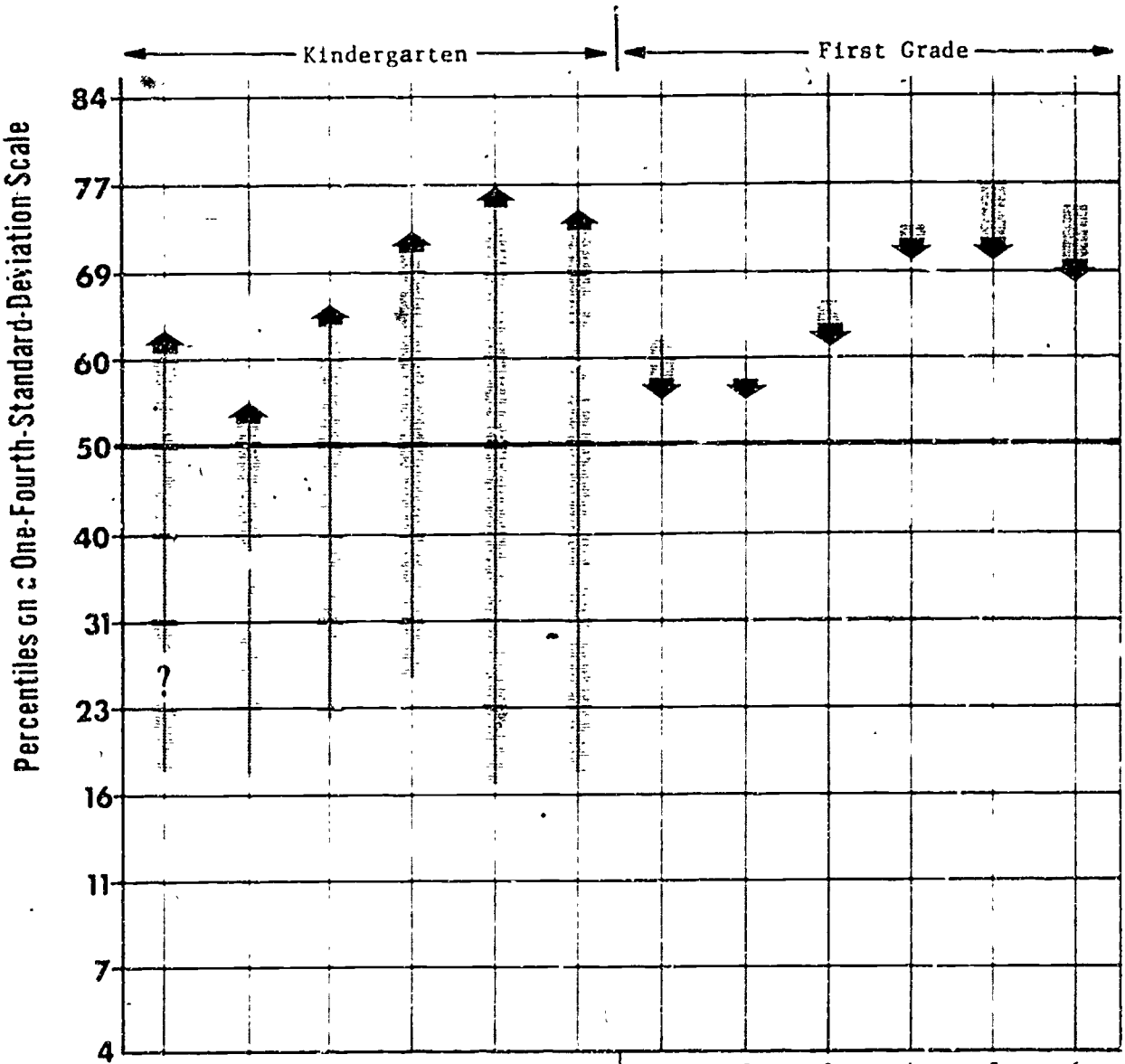
- - - - - - - - -

1st-Starting Sites

The arithmetic data from 1st-starting sites shows an improvement over cohorts of more than one-fourth standard deviation by the end of third grade. The superiority of cohort 1 at the end of kindergarten was not maintained for second and third grade. These findings are consistent with the program modifications that were made.

Table 4.5

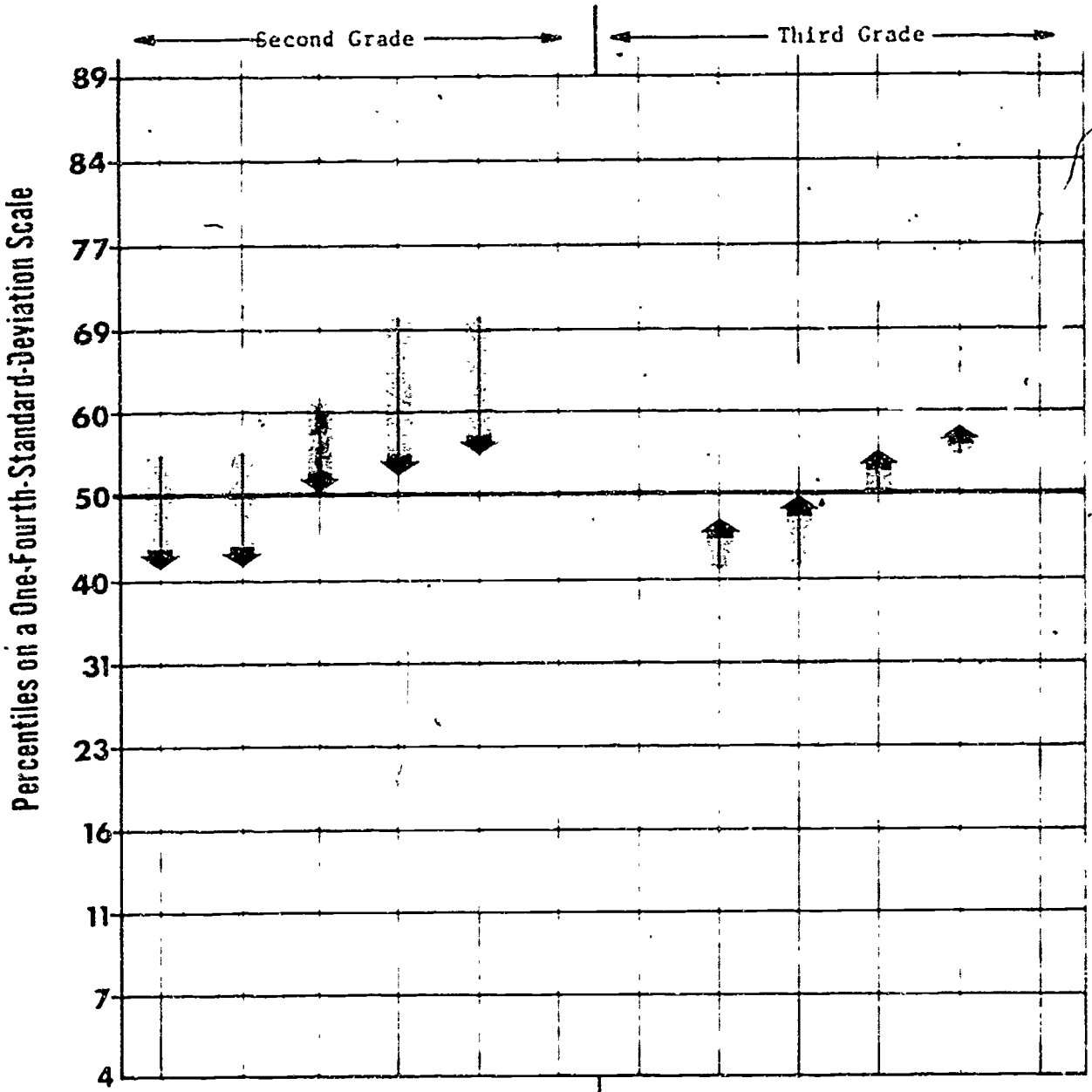
Norm-Referenced Gains (Losses) on WRAT Arithmetic by Cohort from Kindergarten through First Grade for K-Starting, Low-Income, E-B Follow Through Students



	Cohort	1	2	3	4	5	6	1	2	3	4	5	6
	PRE	K	K	K	K	K	K	-	-	-	-	-	-
	POST	K	K	K	K	K	K	K-1	K-1	K-1	K-1	K-1	K-1
Post	N	99	757	1185	1172	1051	925	279	760	1135	992	889	754
Post	%-tile	63	55	66	73	77	75	55	55	61	70	70	68
Post	S. S.	105	102	106	109	111	110	102	102	104	108	108	107
Post	S. D.	14.8	13.5	14.3	14.6	15.4	15.0	10.8	11.4	10.9	12.1	11.4	12.1
Post	G. E.	1.25	1.16	1.32	1.42	1.51	1.43	2.05	2.11	2.19	2.40	2.44	2.34

Table 4.6

Norm-Referenced Gains (Losses) on WRAT Arithmetic by Cohort from Post-First Grade to Post-Third Grade for K-Starting, Low-Income, E-B Follow Through Students



Cohort	1	2	3	4	5	1	2	3	4
POST	1-2	1-2	1-2	1-2	1-2	2-3	2-3	2-3	2-3
Post N	233	656	796	853	750	197	408	669	715
Post %-tile	42	42	50	53	55	47	50	55	58
Post S. S.	97	97	100	101	102	99	100	102	103
Post S. D.	6.5	8.5	9.3	9.8	9.2	8.6	8.5	8.5	8.7
Post G. E.	2.7[illegible]	2.82	3.07	3.11	3.23	3.70	3.84	4.06	4.13

The gain against the norm group is consistently shown to occur in first and third grades, not second. These gains are consistent with the findings for K-starting sites.

- - - - - - - - - - -

Tables 4.7 and 4.8

- - - - - - - - - - -

MAT

An improvement over cohorts in arithmetic skills is shown on the MAT. For K-starting sites at the end of third grade in cohorts 2, 3, and 4, the improvement on Total Math is from a mean standard score of 66.3 to 72.6 to 72.7. In percentiles, this is an increase from 35 to 59 or one-half standard deviation unit. For 1st-starting sites at the end of third grade in cohorts 3, 4, and 5, the improvement on Total Math is from a mean standard score of 68.7 to 70.7 to 71.2. In percentiles, this is an increase from 41 to 52. This is a one-quarter standard deviation improvement.

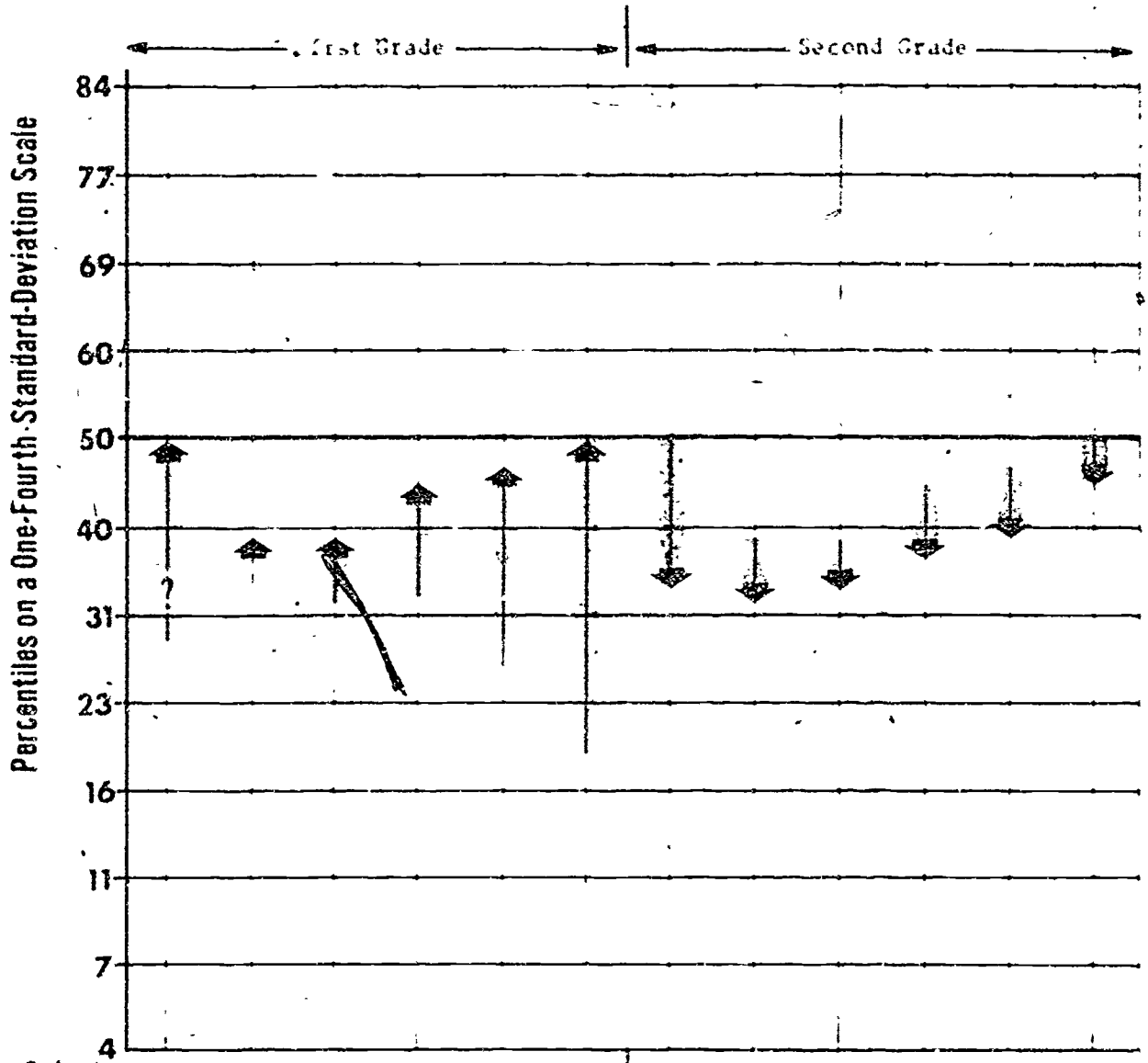
SPELLING

K-Starting Sites

In many ways, spelling was neglected by the sponsor in the process of giving priority to reading, oral communication skills, and logical and mathematical competence. The local projects were encouraged to provide their own spelling programs. Some did, but many did not. After several years of failure to implement spelling programs, our project managers were encouraged to be sure spelling was included at all levels in all sites. The effects of this effort are particularly apparent on the WRAT Spelling test which was given to all cohorts. Tables

Table 4.7

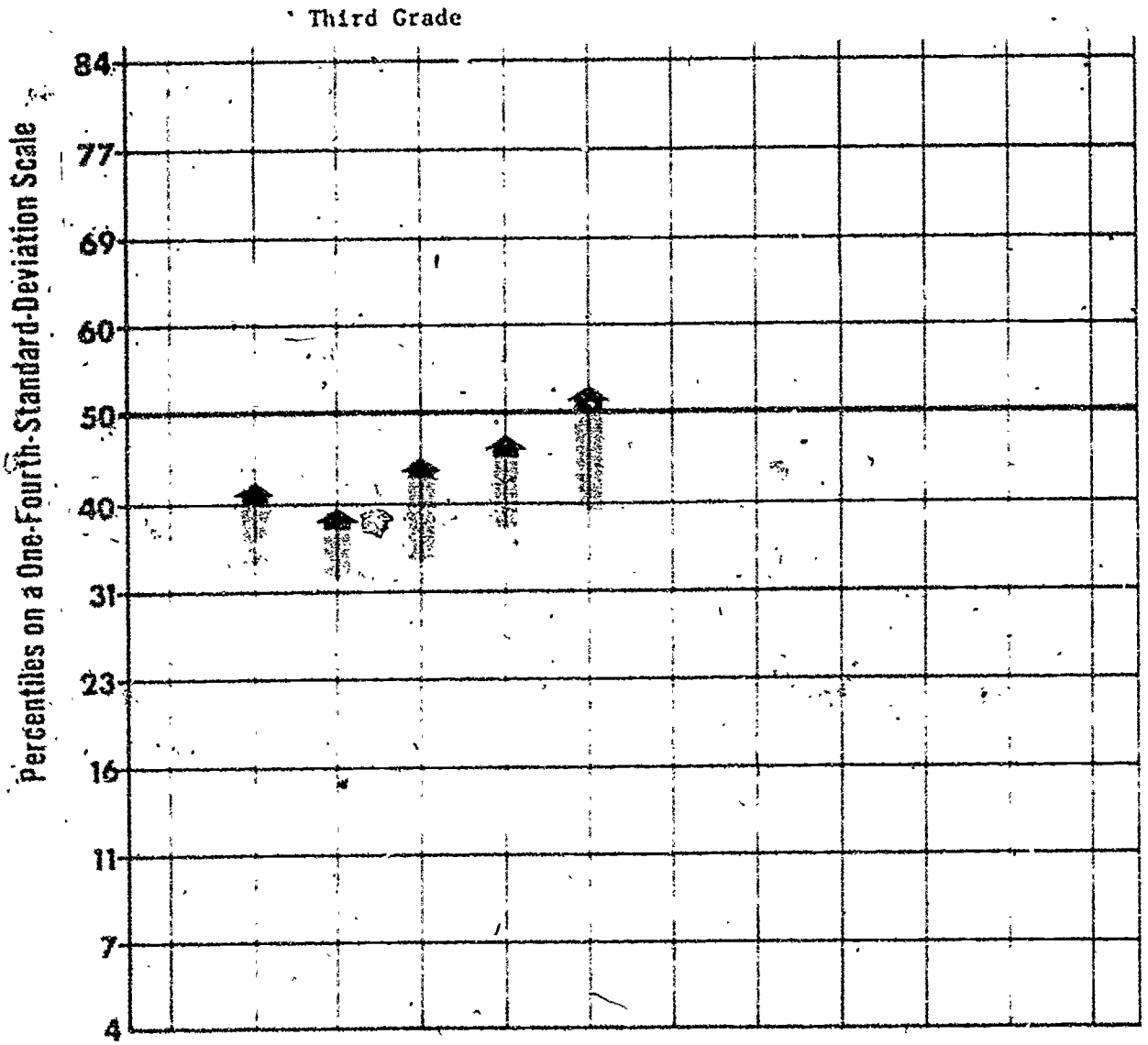
Norm-Referenced Gains on WRAT Arithmetic by Cohort
from Pre-First Grade to Post-Second Grade for
1st-Starting, Low-Income, E-B Follow Through Students



	Cohort	1	2	3	4	5	6	1	2	3	4	5	6
	PRE	1	1	1	1	1	1	-	-	-	-	-	-
	POST	1	1	1	1	1	1	1-2	1-2	1-2	1-2	1-2	1-2
Post	N	21[illegible]	1[illegible]	1[illegible]	[illegible]	711	703	222	1162	479	[illegible]	606	541
Post	%-tile	[illegible]	[illegible]	34	47	7	50	34	32	34	47	39	45
Post	S. S	1[illegible]	[illegible]	[illegible]	[illegible]	94	100	[illegible]	[illegible]	[illegible]	[illegible]	98	98
Post	S. D.	12.1	12.2	11.[illegible]	[illegible]	11.1	11.2	6.1	6.2	9	9.6	8.4	10.7
	G. E.	2.06	1.[illegible]9	1.78	1.[illegible]9	1.97	2.04	2.65	2.55	2.64	2.77	2.82	2.93

Table 4.8

Norm-Referenced Gains on WRAT-Arithmetic by Cohort from Post-Second Grade to Post-Third Grade for 1st-Starting, Low-Income, E-B Follow Through Students



Cohort	1	2	3	4	5
POST	2-3	2-3	2-3	2-3	2-3
Post N	677	1012	852	602	484
Post %-tile	42	39	45	47	53
Post S. S.	97	96	98	99	101
Post S. D.	11.2	9.4	9.8	9.6	8.5
t G. E.	3.69	3.57	3.81	4.02	4.03

4.9 and 4.10 show the data for each grade level. The improvement over cohorts is nearly one-standard deviation in kindergarten. The magnitude of improvement then decreases in successive grade levels, so that by the end of third grade, no significant effect is apparent (with or without projections for cohorts 5 and 6). Better ways of teaching advanced spelling skills are needed and are currently a programmatic concern of the sponsor.

A consistency of gains against the norm group is present over cohorts. Substantial gains are made in kindergarten (more than one-and one-half standard deviation units for some cohorts) and then there is a loss of gains, dropping back to the 50th percentile.

- - - - - - - - - - -

Tables 4.9 and 4.10

- - - - - - - - - - -

1st-Starting Sites

Tables 4.11 and 4.12 show the improvements over cohorts for 1st-starting sites. Small trends for improvement are shown, about one-fourth standard deviation in magnitude.

- - - - - - - - - - -

Tables 4.11 and 4.12

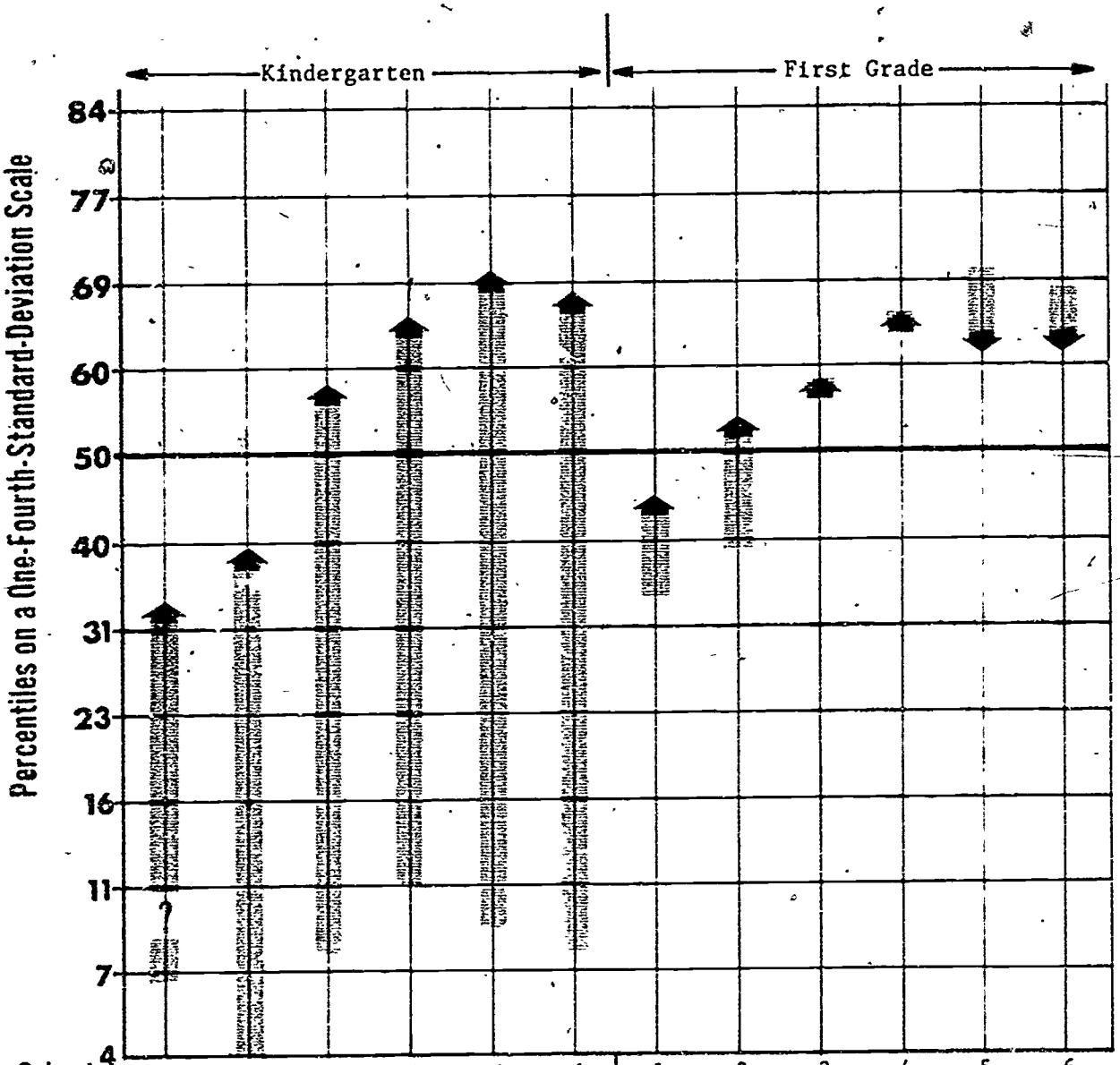
- - - - - - - - - - -

MAT

MAT Spelling scores from K-sites at the end of third grade show no systematic improvements over cohorts 2, 3, and 4. For 1st-sites an improvement of less than one-fourth standard deviation is projected.

Table 4.9

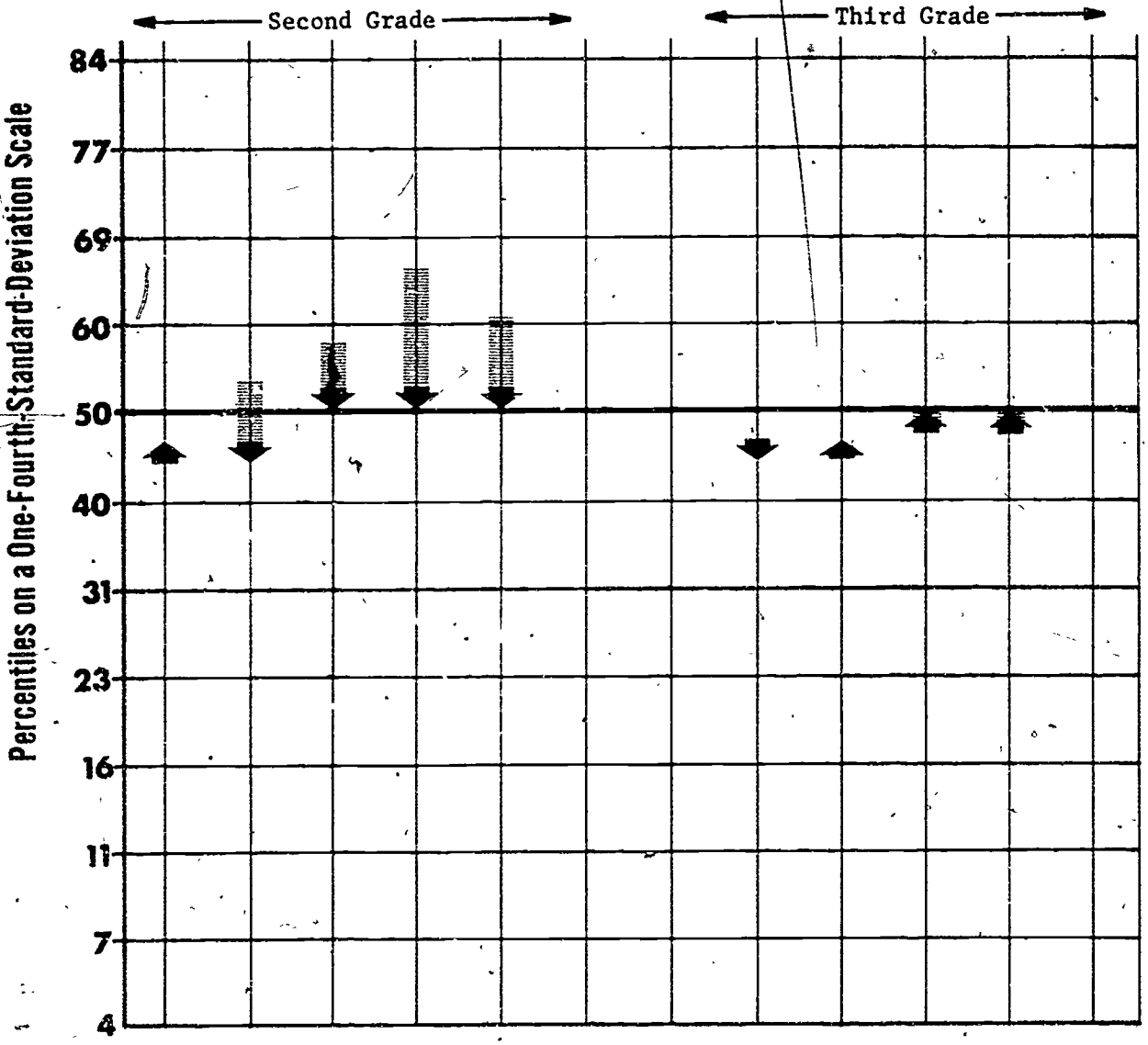
Norm-Referenced Gains (Losses) on WRAT Spelling by Cohort from Kindergarten through First Grade for K-Starting, Low-Income, E-B Follow Through Students



Cohort	1	2	3	4	5	6	1	2	3	4	5	6
PRE	K	K	K	K	K	K	–	–	–	–	–	–
POST	K	K	K	K	K	K	K-1	K-1	K-1	K-1	K-1	K-1
N	99	764	1185	1203	1053	930	283	763	1137	992	891	764
%-tile	34	39	58	66	70	68	45	53	58	66	61	61
S. S.	94	96	103	106	108	107	98	101	103	106	104	104
S. D.	13.0	15.9	15.1	15.5	16.6	16.1	12.6	13.0	13.1	14.1	13.2	14.5
G. E.	.82	.88	1.19	1.30	1.38	1.31	1.82	2.05	2.16	2.29	2.22	2.18

Table 4.10

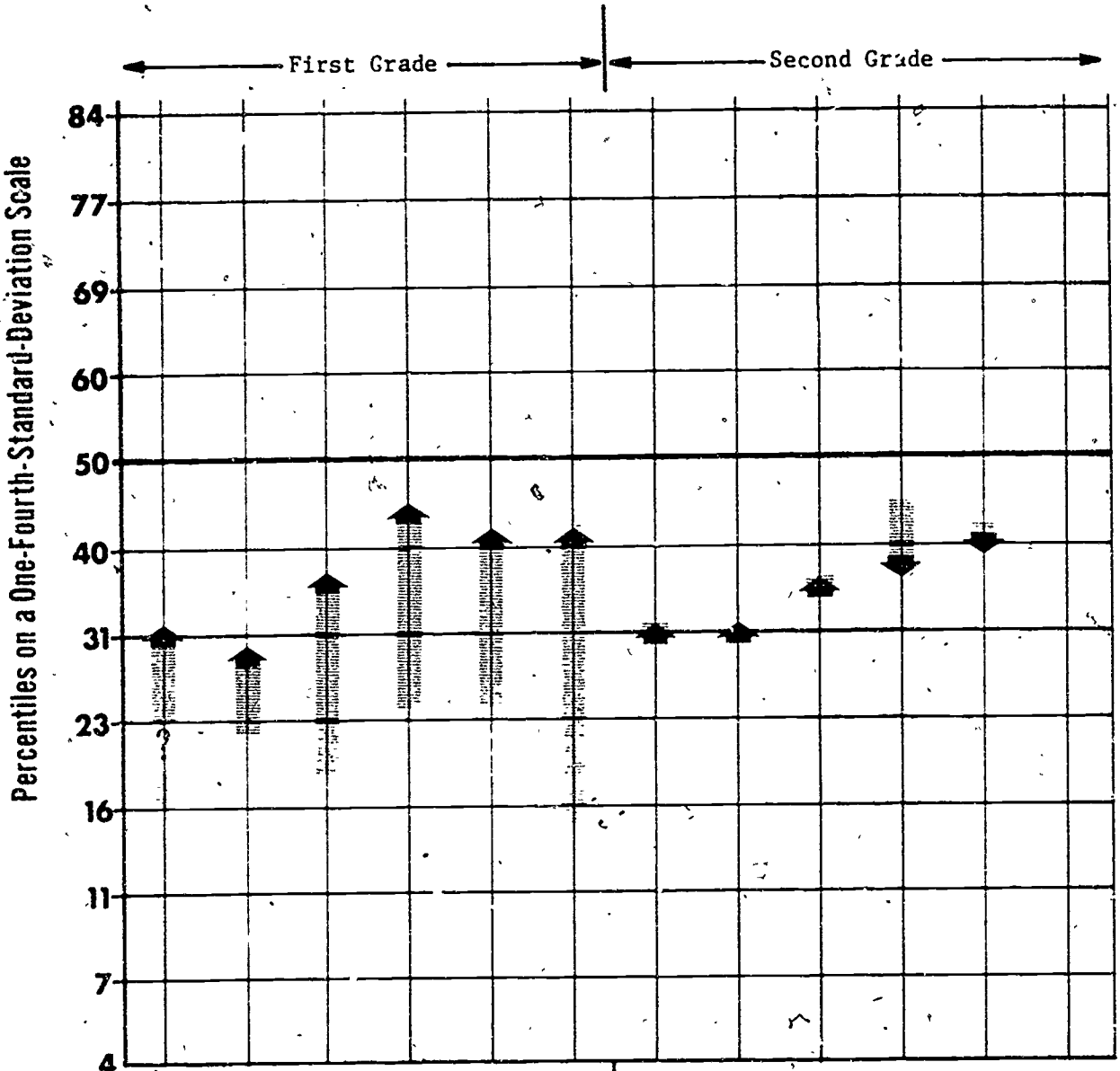
Norm-Referenced Gains (Losses) on WRAT Spelling by Cohort from Post-First to Post-Third Grade for K-Starting, Low-Income, E-B Follow Through Students



Cohort	1	2	3	4	5	1	2	3	4
POST	1-2	1-2	1-2	1-2	1-2	2-3	2-3	2-3	2-3
Post N	234	656	796	854	758	199	408	649	718
Post %-tile	47	45	50	50	50	45	47	50	50
Post S. S.	99	98	100	100	100	98	99	100	100
Post S. D.	11.2	13.7	13.8	13.6	13.9	12.2	13.2	13.4	15.0
G. E.	2.92	2.91	3.06	3.04	3.06	3.56	3.73	3.90	3.91

Table 4.12

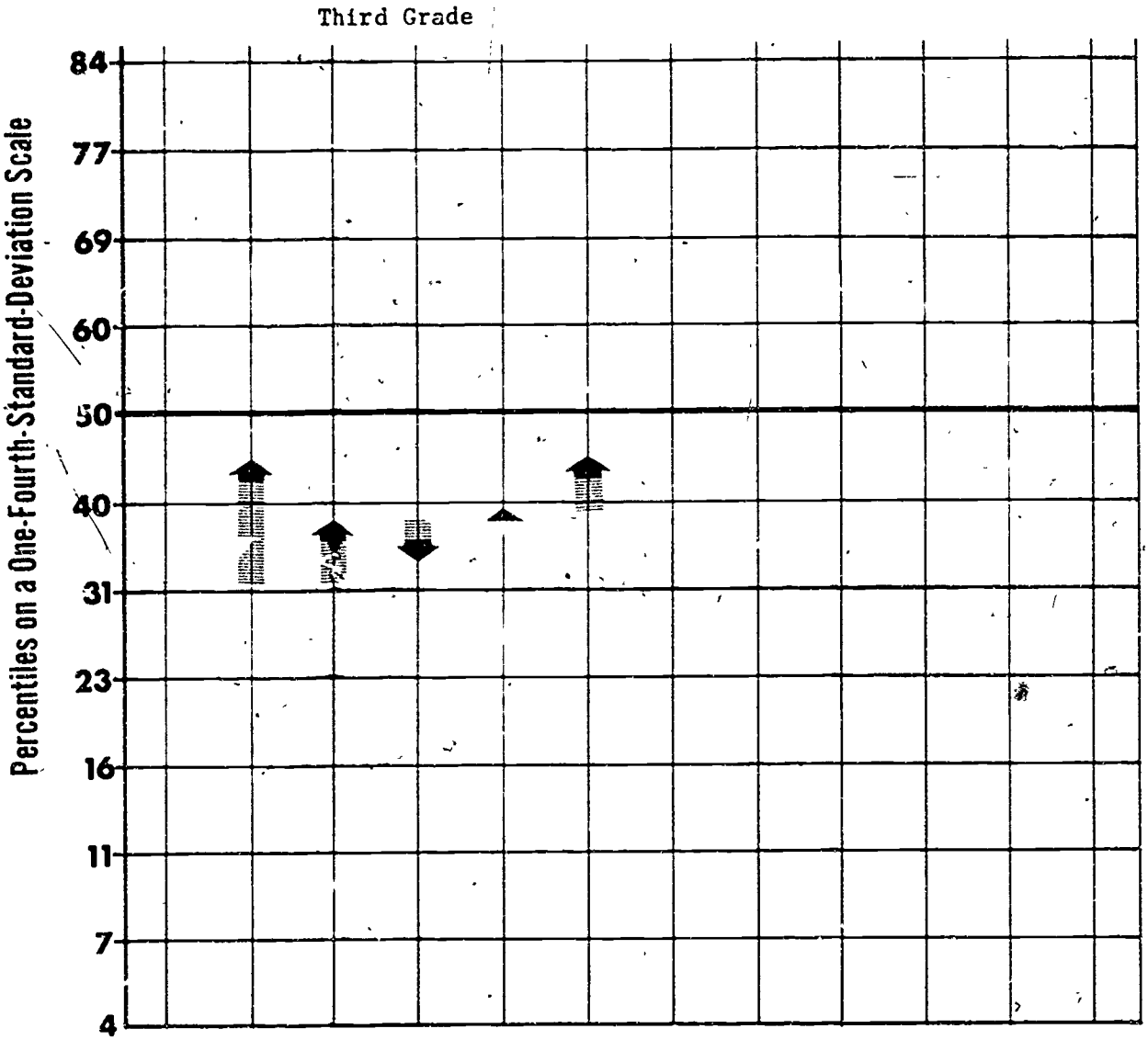
Norm-Referenced Gains on WRAT Spelling by Cohort from Pre-First Grade to Post-Second Grade for 1st-Starting, Low-Income, E-B Follow Through Students



Cohort	1	2	3	4	5	6	1	2	3	4	5
PRE	1	1	1	1	1	1	-	-	-	-	
POST	1	1	1	1	1	1	1-2	1-2	1-2	1-2	1-2
Post N	727	1216	1014	740	712	764	778	1163	948	606	610
Post %-tile	32	30	37	45	42	42	32	32	37	37	39
Post S. S.	93	92	95	98	97	97	93	93	95	95	96
Post S. D.	13.1	12.6	12.2	14.1	12.7	13.3	12.4	12.0	13.1	13.5	11.2
Post G. E.	1.66	1.56	1.73	1.90	1.87	1.88	2.55	2.53	2.67	2.72	2.79

Table 4.12

Norm-Referenced Gains on WRAT Spelling by Cohort
from Post-Second Grade to Post-Third Grade for
1st-Starting, Low-Income, E-B Follow Through Students



Cohort	1	2	3	4	5
POST	2-3	2-3	2-3	2-3	2-3
Post N	677	1013	855	576	485
Post %-tile	45	37	34	39	45
Post S. S.	98	95	94	96	98
Post S. D.	18.1	14.4	13.4	14.8	14.8
Post G. E.	3.80	3.48	3.40	3.63	3.79

IQ GAINS

An examination of IQ gains using the SIT over cohorts will not be undertaken for K-starting sites because of the limited data in the PRE-POST GAINS analysis. For the 1st-starting sites, the gains data are shown by cohort in Table 4.13. The cohort 1 data is very limited (based on only one site where the entry level of the children is quite low). However, this cohort shows a level of gains that was not shown again until cohort 5. From cohort 2 to cohort 5 some improvement in gain is to be noted. The difference in gain by cohort 3 over cohort 2 and by cohort 5 over cohort 4 are significant beyond the .001 level. The size of the gains appear to be correlated with entry level IQ.

- - - - - -

Table 4.13

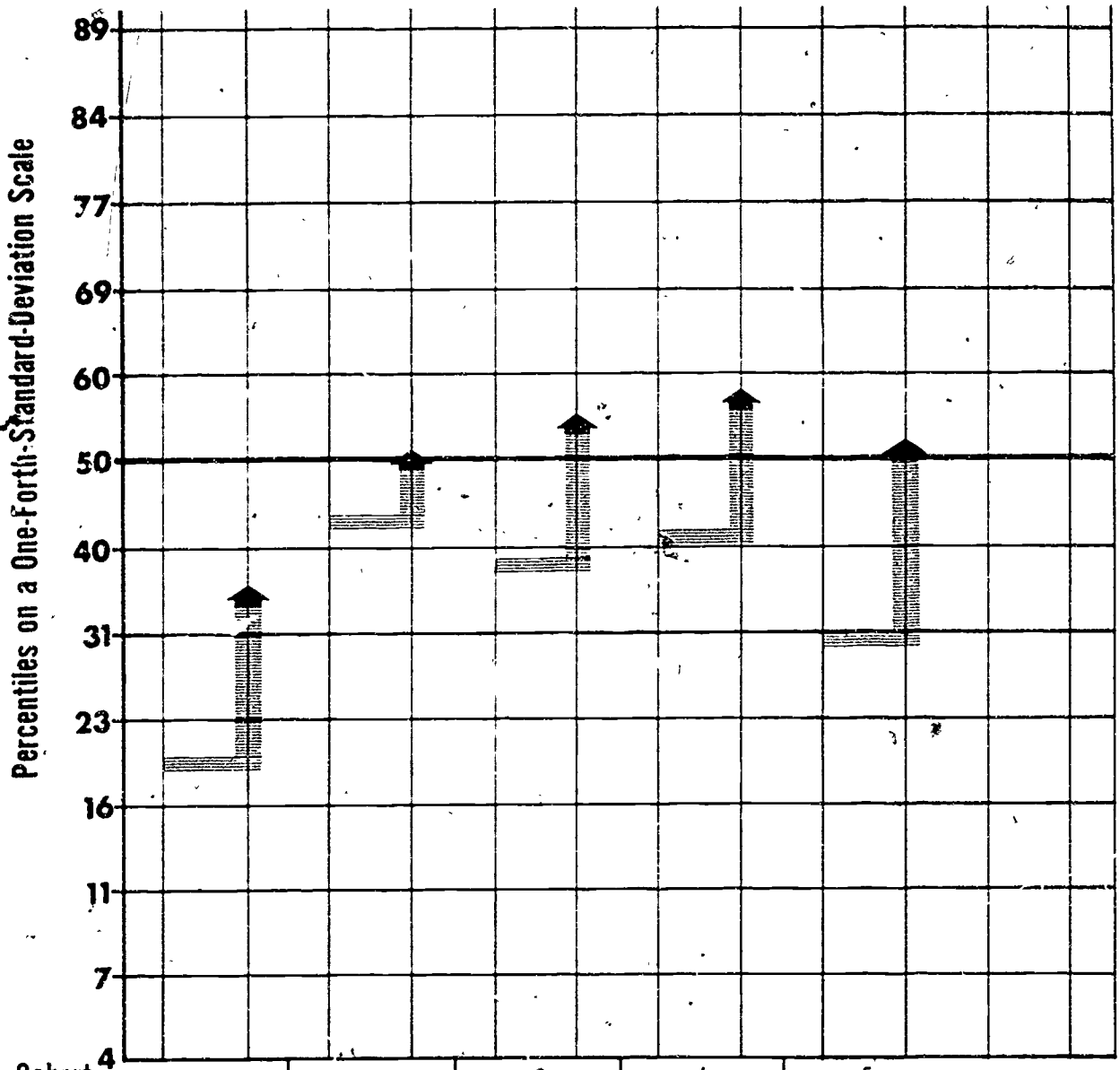
- - - - - -

Summary

The analyses by cohorts show that the cohort 1 (though based on limited data) was somewhat superior in progress to the next two. The program changes introduce to insure that all low performers were taught may have actually slowed down some better students and led to lower mean outcome levels. Nevertheless, cohorts 2 to 6 show a progressive improvement in performance in K-sites on WRAT Reading and Arithmetic (and to a lesser degree Spelling). Some progressive improvements on the MAT are also shown for the cohorts tested with the MAT. The analysis revealed that the major source of the improvement over cohorts was associated with improvements in implementation at the kindergarten level. It took some time to get kindergarten teachers to make more serious efforts to teach academics to young children. Many had been trained that it was quite inappropriate to do so.

Table 4.13

SIT IQ Gains over Cohorts for 1st-Starting, Low-Income, E-B Follow Through Students from the PRE-POST GAINS Analysis



Cohort	1		2		3		4		5	
PRE	1	-	1	-	1	-	1	-	1	-
POST	-	3	-	3	-	3	-	3	-	3
N	57	57	210	210	311	311	136	136	282	282
%-tile	19.2	35.1	43.2	50.6	38.0	54.8	41.7	58.0	30.2	52.4
S. S.	87.1	94.4	97.4	100.2	95.5	101.9	96.9	103.0	92.1	100.8
S. D.	13.1	22.2	15.9	16.7	14.9	16.2	13.1	14.6	12.3	16.2
Gain	7.35		2.76		6.42		6.05		8.73	
D. Gain	18.9		15.2		14.3		13.1		14.1	

In keeping with this, less pronounced improvements over cohorts 2 to 6 were found for 1st-sites. The largest improvement was found in arithmetic, where successive program modifications may account for the gains.

The PRE-POST GAINS analysis was used to examine IQ gains for 1st-starting sites only. Discounting the cohort 1 which showed a nice gain (7.35) for 57 children in one site, the improvement gains from cohort 2 to 5 was from 2.76 to 8.73.

CHAPTER 5

CONTROLLED WITHIN-SITE COMPARISONS (INCLUDING HEAD START)

In a number of our projects we started at several grade levels at once, or the sites switched from being 1st-starting to K-starting after Follow Through was under way. Where these events happened, it is possible to make comparisons within the same schools with the same teachers for children who have had two years of the E-B program, three years of the program, or four years of the program. If the program is having an impact, those children who are in it longer should do better. These comparisons also provide another kind of answer to the question "Does kindergarten make a difference?" Since the comparisons of K-starting sites with 1st-starting sites are possibly confounded by regional differences in socio-economic conditions, these within-site comparisons permit an assessment of the possible gains to be made with the disadvantaged by a fuller use of kindergarten for basic instruction.

In this chapter we will also examine data for three sites where Planned Variation Head Start was introduced in 1969, 1970, and/or 1971. Examination of the progress of kids with and without a sponsored Head Start program is possible. In addition, it is possible to evaluate the effects of sponsored Headstart programs for five-year-olds versus sponsored public school kindergarten for five-year-olds.

CHEROKEE

The first within-site comparisons come from the Eastern Band Cherokee Indian community in North Carolina where the children attend a BIA school. In 1970, we installed our program simultaneously in kindergarten, first, and second grades. The children had had the benefit of Follow Through the year before with another sponsor who was terminated. As far as we could determine, there had been no significant impact of that year of Follow Through on the school program. Tables 5.1 to 5.5 provide the comparisons on WRAT and MAT variables for the children who started the program in second grade (2 years of E-B), first grade (3 years of E-B) and kindergarten (4 years of E-B). The K-starting group includes four cohorts through first grade, three through second, and two through third. Examination of the K-starting data by cohorts shows progressive improvements for the first three cohorts.

Table 5.1 gives the data for WRAT Reading. It can be seen that the children with 4 years of program are nearly a standard deviation above the children with 2 years of program. The difference is highly significant ($p < .001$). Children with 4 years of program are also well ahead of those with only 3 years of program at the end of first and second grades, but not third grade. Apparently 3 years was enough to catch up on decoding skills.

- - - - - -

Table 5.1

Table 5.2

Table 5.3

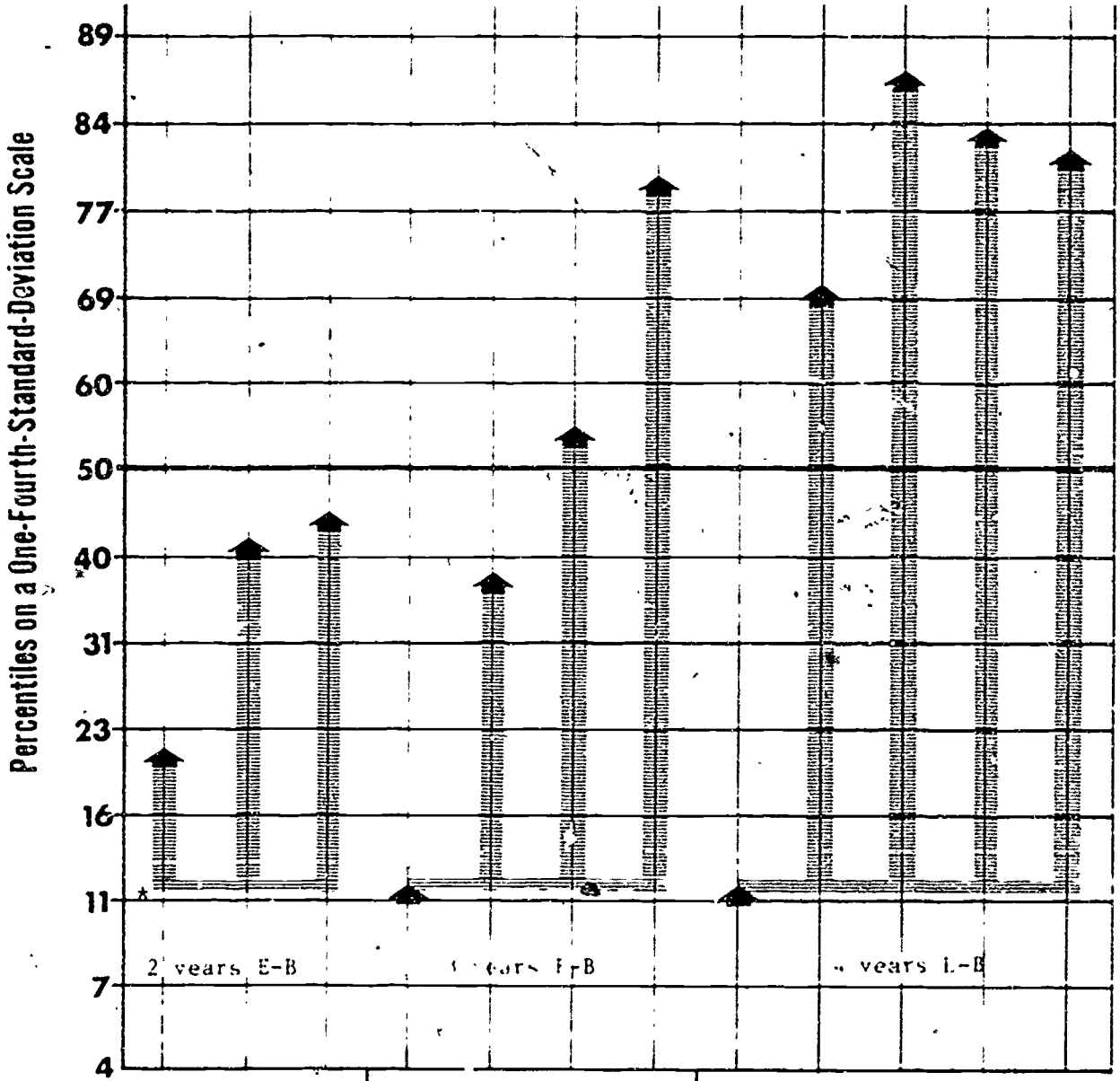
Table 5.4

Table 5.5

- - - - -

Table 5.1

Comparison of Progress for All Cherokee Students with 2, 3, and 4 Years of E-B Follow Through on WRAT Reading
(Retentions with Agemates)

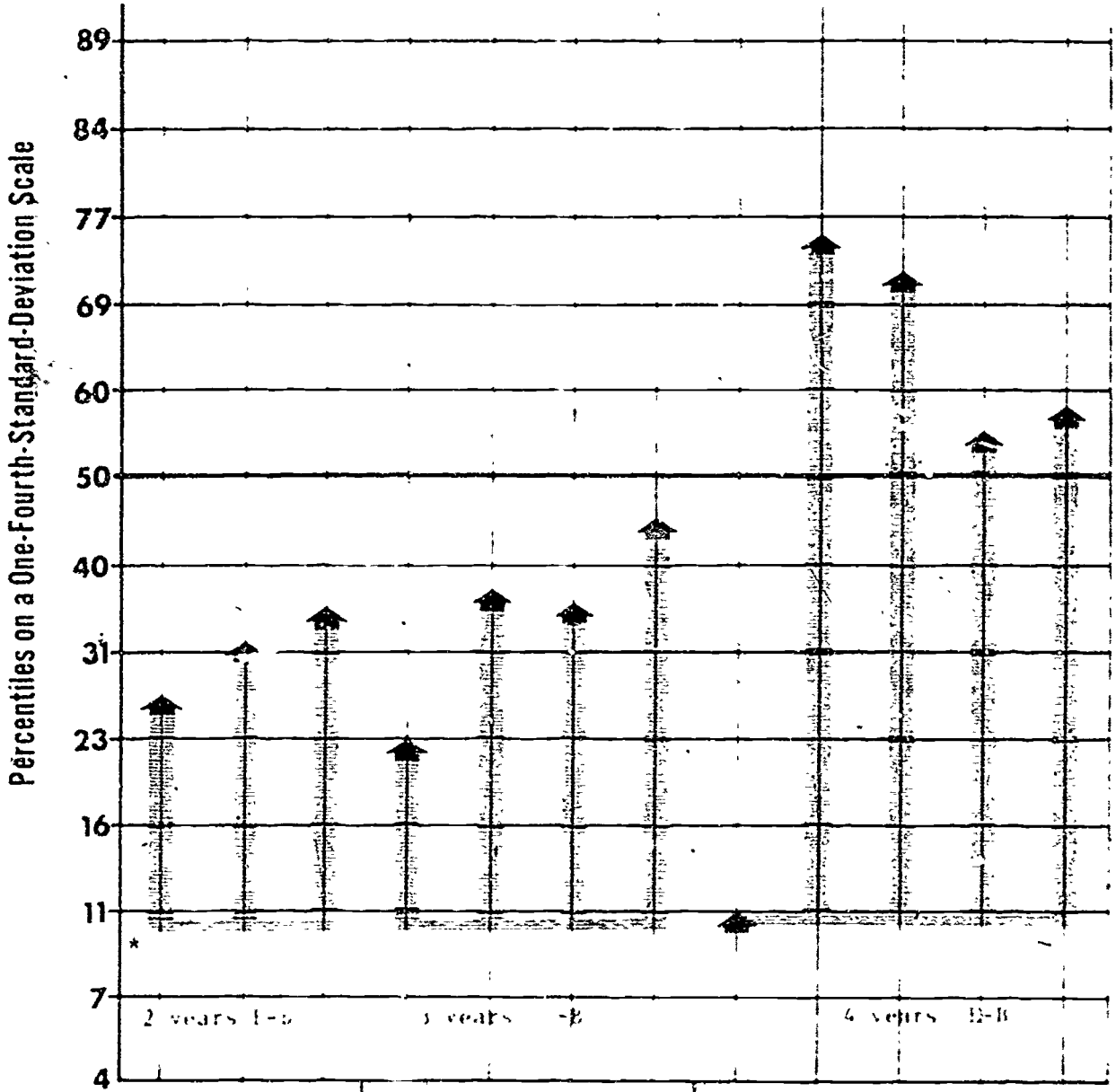


PRE	-	-	-	-	-	-	-	-	-	-	-	-
POST	[illegible]											
N	66	[illegible]	100	[illegible]								188
%-tile	21	42	[illegible]	[illegible]								
S. S.	[illegible]	[illegible]	[illegible]									
S. D.	[illegible]	[illegible]	[illegible]									
G. E.	1.69	[illegible]	[illegible]									

*Four-year pre-K baseline used for all groups.

Table 5.2

Comparison of Progress for All Cherokee Students with 2, 3, and 4 Years of E-B Follow Through on WRAT Arithmetic
(Retentions with Agemates)

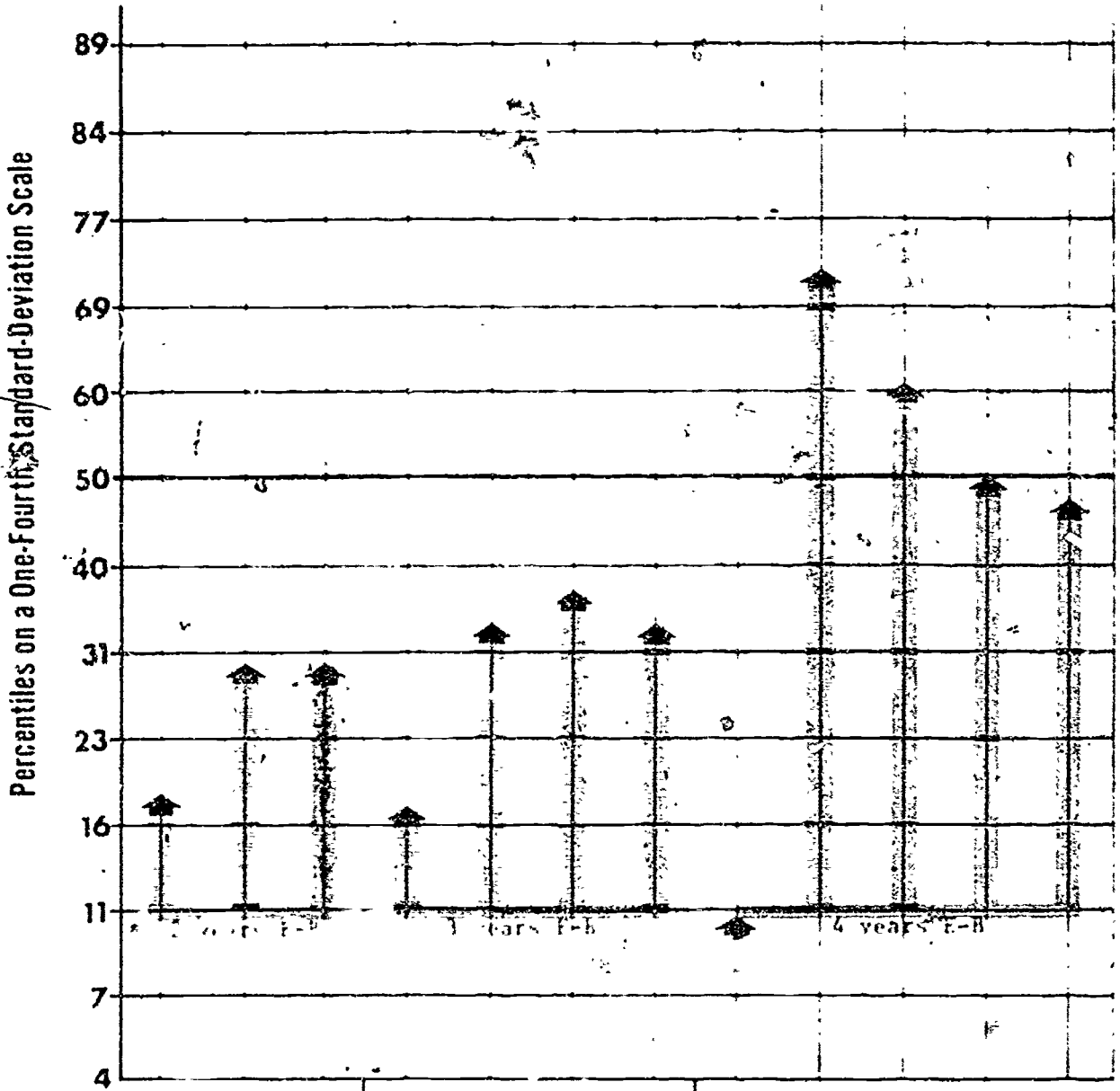


PRE												
POST												
N												
%-tile												
S. S.												
S. D.												
G. E.												

*Four-year pre-f baseline used for all groups.

Table 5.3

Comparison of Progress for All Cherokee Students with 2, 3, and 4 Years of E-8 Follow Through on WRAT Spelling
(Retentions with Agemates)

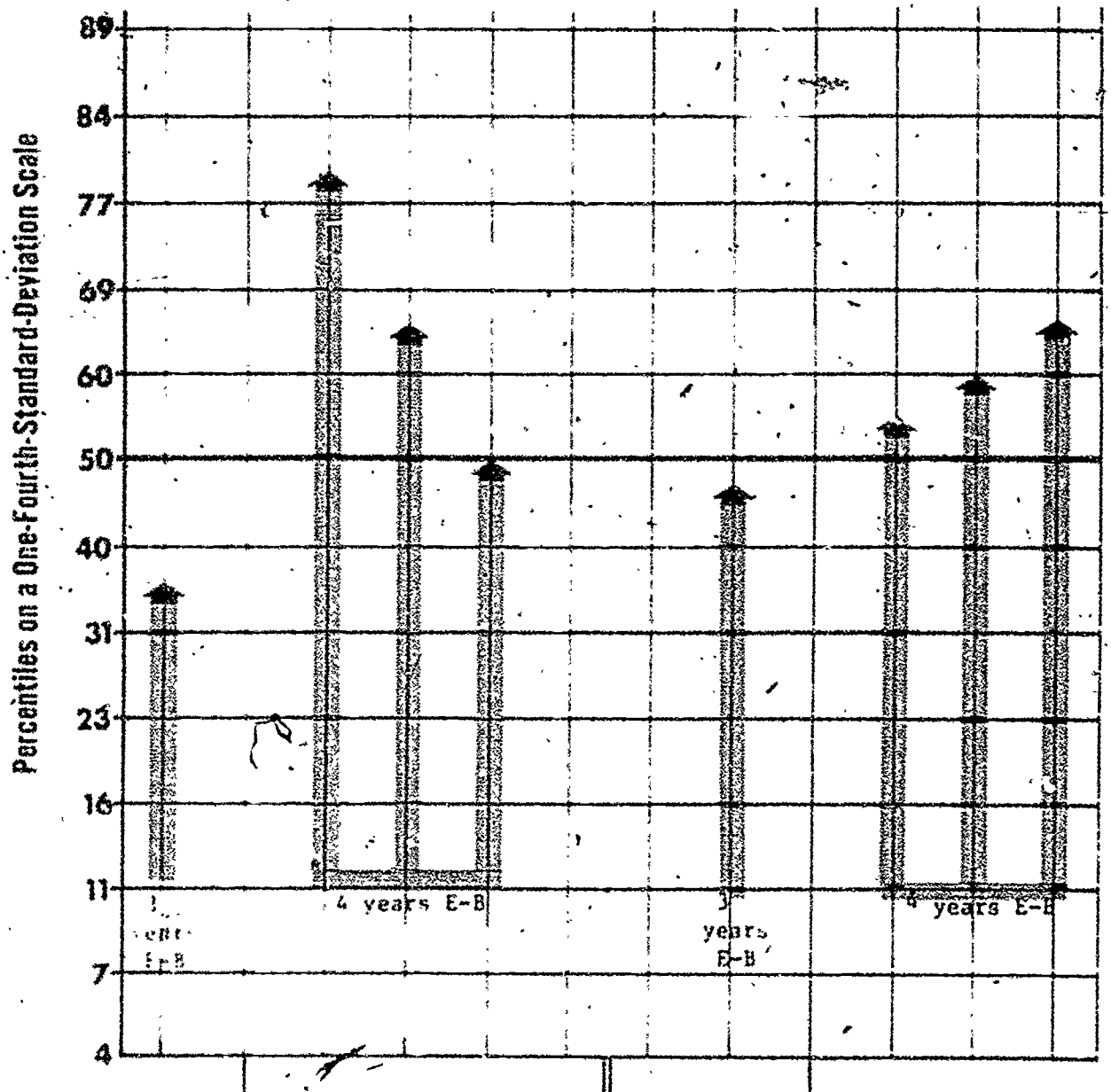


PRE			
POST			
N			
%-tile			
S S			
S.D.			
G.E			

*Four-year pre-test baseline used for all groups.

Table 5.4

Comparison of Progress for All Cherokee Students with 3 and 4 Years of E-B Follow Through on MAT

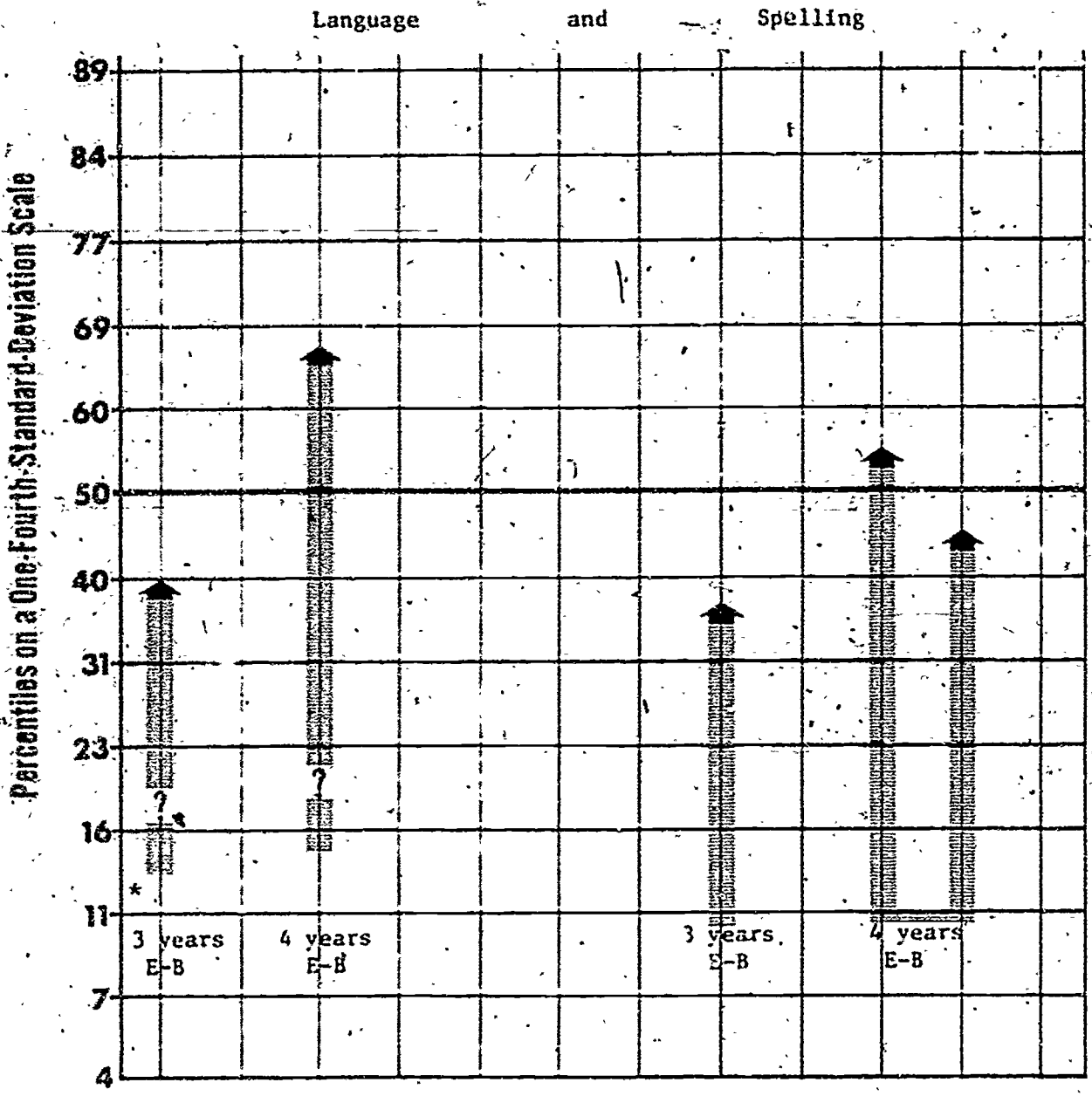


POST				3	1	2	3	
N	109	[illegible]	263	185	107	244	261	185
%-tile	[illegible]	[illegible]	[illegible]	[illegible]	47	55	60	66
S. S.	[illegible]	[illegible]	[illegible]	65.9	69.7	44.7	59.4	75.1
S. D.	[illegible]	[illegible]	9.7	10.8	9.8	11.3	11.0	12.2
G. E.	3.20	[illegible]	3.08*	3.61	3.80	2.03	2.96	4.37

* WRAT four-year baselines used.

Table 5.5

Comparison of Progress for All Cherokee Students with 3 and 4 Years of E-B Follow Through on MAT



POST	3	3	3	2	3
N	104	167	103	246	184
%-tile	40	67	36	54	45
S. S.	65.8	75.7	59.9	57.6	62.3
S. D.	10.7	12.2	11.2	10.7	12.3
G. E.	3.69	5.03	3.27	3.11	3.64

* WRAT four-year baselines used.

As Table 5.2 shows, the 4-year students did significantly better ($p < .001$) than both the 3-year group and the 2-year groups on WRAT Arithmetic. Table 5.3 shows that this same conclusion holds for WRAT Spelling. The magnitude of the differences at the end of third grade are from a quarter to a half standard deviation.

The comparisons for MAT Total Reading and Total Math (Table 5.4) also show significant effects favoring the 4-year group. The difference for MAT Total Reading is significant at the .002 level, and that for Total Math is beyond the .001 level. Table 5.5 shows the effects for Language and Spelling on the MAT. The Language difference significantly favors the 4-year group ($p < .001$) as does the Spelling difference ($p < .02$).

IQ data are available for comparing the 3-year group with the 4-year group. The comparisons are based only on the students for whom there is a pretest and posttest. The data are as follows:

	Pre-1			Post 3			Gain		
	Mean	SD	N	Mean	SD	N	Mean	SD	N
3-year group	98.6	17.3	75	101.7	18.6	75	3.05	12.4	75
	Pre-K			Post 3			Gain		
4-year group 1	101.4	10.0	49	107.9	15.5	49	9.38	13.5	73
4-year group 2	100.5	13.7	24	115.7	13.9	24			

The difference in IQ gain is significant at the .01 level. The difference at the end of third grade is also significant.

These comparisons clearly show a strong program impact that increased the longer the program was in effect.

E. ST. LOUIS

E. St. Louis, Illinois is a Black community just over the river from St. Louis. E. St. Louis started Follow Through in 1968 as a 1st-grade-starting parent-implemented program who chose the E-B model as sponsor. In 1970, public school kindergartens were introduced. This gave us three cohorts with 1st-starting children and four with K-starting children. In addition, Planned Variation Head Start was introduced as a program for five-year olds in 1969 and for four-year olds in 1970. The Head Start data will be discussed later in this chapter. For now, the children with E-B Sponsored Head Start are excluded from the analysis.

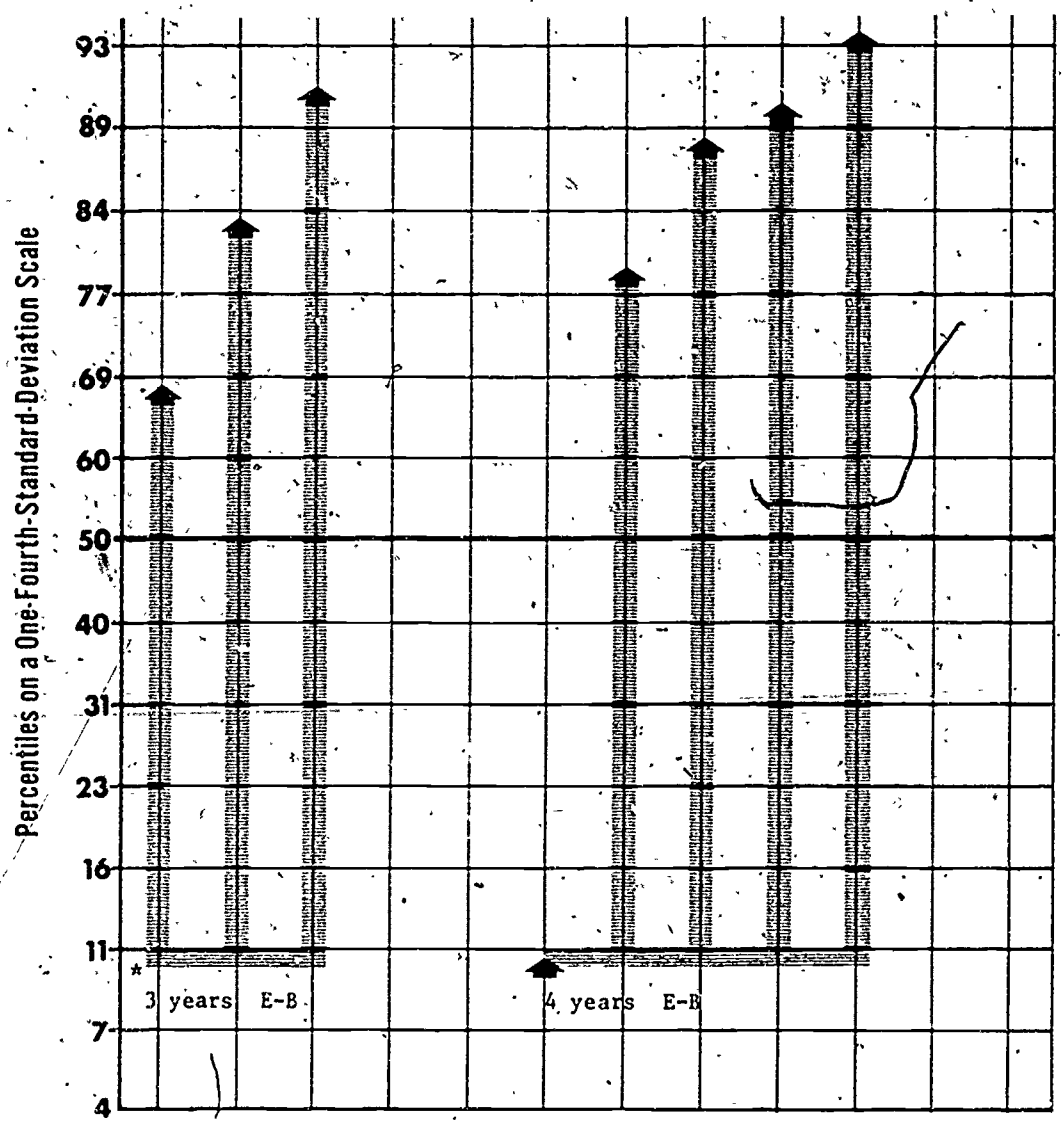
It was noted in discussing the data for 1st-starting sites over cohorts, that cohort 1 for E. St. Louis performed unusually high on most measures at the end of third grade. They had been taught by a selected group of teachers and made great progress. This unusual performance for cohort 1 tends to obscure some of the effects of K-starting and 1st-starting cohorts. Tables 5.6 to 5.10 present the comparisons for children with 3 and 4-years of program. For WRAT Reading, the differences are significant and favor K-starting children at the end of grade 1 and 2, but not 3. The unusually, high performance of cohort 1 children at the end of grade 3 reduced the difference. Separated by cohort the means at the end of third grade are:

	1st-Starting (3-years E-B)			
	G.E.	SS	SD	N
Cohort 1	7.28	129	25.8	166
Cohort 2	5.34	115	21.8	122
Cohort 3	4.91	110	20.8	98
	K-Starting (4-years E-B)			
Cohort 3	6.37	125	24.1	89
Cohort 4	6.17	121	26.8	71

Table 5.6

Table 5.6

Comparison of Progress for Low-Income E. St. Louis Students with 3 and 4 Years of E-B Follow Through on WRAT Reading (Retentions with Grademates)



PRE	-	-	-	K	-	-	-	-
POST	1	2	3	-	K	1	2	3
N	431	419	389	238	610	610	371	160
%-tile	68	83	91	10	79	88	90	94
S. S.	107	114	120	81	112	118	119	123
S. D.	21.0	21.3	23.5	18.8	18.3	22.2	22.9	25.4
G. E.	2.38	4.20	6.05	-.03	1.49	3.06	3.72	6.28

*Four-year pre-K baseline.

Cohorts 5 and 6 will be stronger than the K-starting cohorts 3 and 4 if current trends continue. Thus, it is apparent that the performance of cohort 1 on WRAT Reading is atypical. If these data were to be excluded from the evaluation, there would be an advantage for the K-starting cohorts of three-fourths of a standard deviation (112 versus 123).

On WRAT Arithmetic, cohort 1 is nearly a grade level ahead of cohorts 2 and 3 (two-thirds standard deviation), but even given this atypical performance, the K-starting cohorts show a significantly superior performance at all grade levels ($p < .001$). Excluding cohort 1, the superiority is about .7 grade levels.

- - - - - - - - - - - - - - -

Tables 5.7, 5.8, 5.9, 5.10

- - - - - - - - - - - - - - -

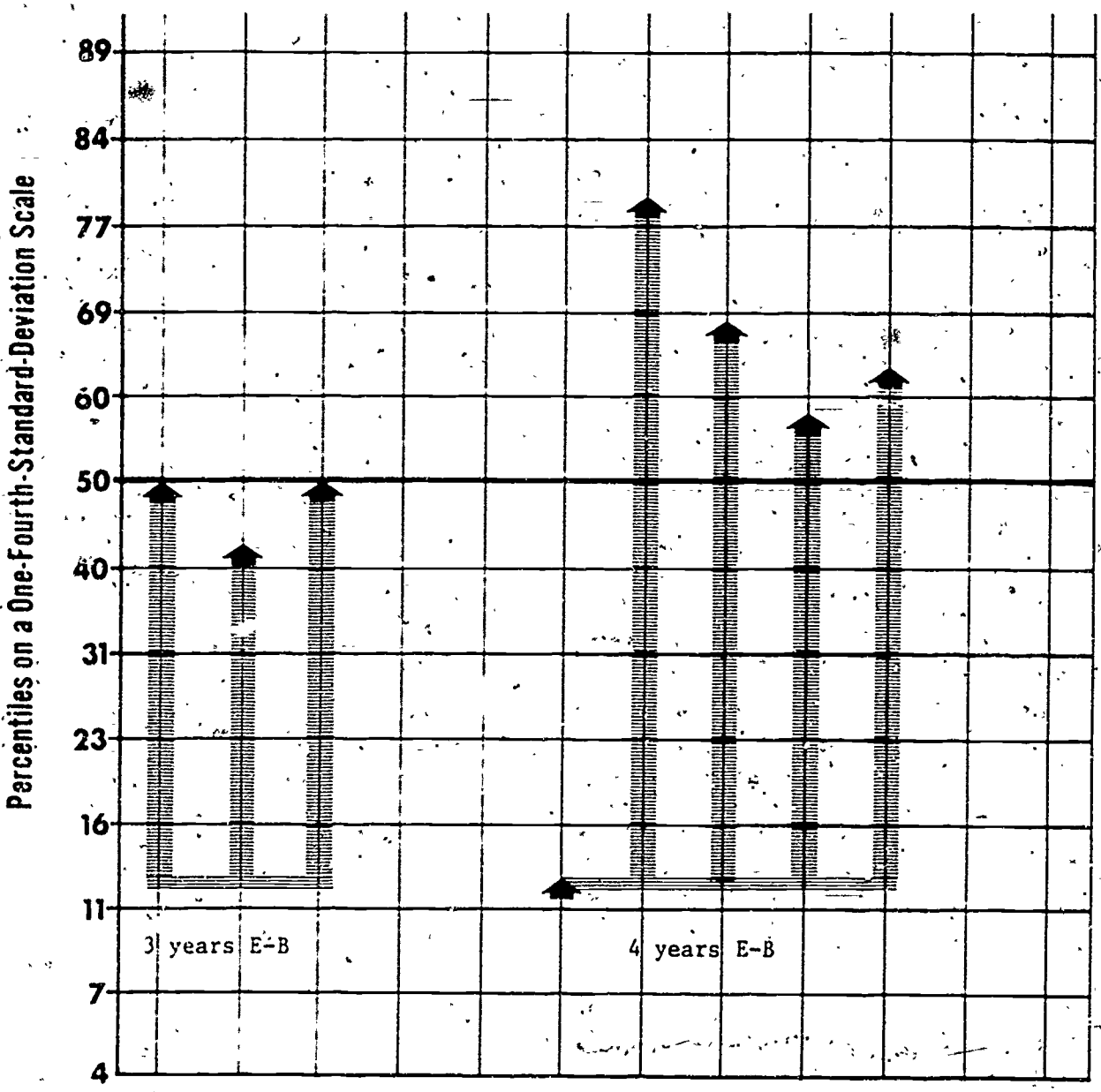
On WRAT Spelling (Table 5.8), again the highly atypical performance of cohort 1 (cohort 1 = 112, cohort 2 = 99, cohort 3 = 94) eliminates any significant effect by the end of third grade. Comparisons excluding cohort 1 are highly significant, as are effects at the end of first and second grades.

On the MAT the comparisons involved 1st-starting cohort 3 and with K-starting cohorts 3 and 4. The differences favor the K-starting groups at the .001 level or better for Total Reading and Total Math (Table 5.9). The magnitude of the differences is about one-half standard deviation or .4 grade levels in each case. On MAT Language, there is no significant difference, while the difference on Spelling favors the K-starting groups at better than the .01 level.

Comparisons on the SIT show more between-year gains on for K-starting than for 1st-starting children, but because the testing with the SIT does not cover enough of the same kids at pretest and posttest, they will not be reported.

Table 5.7

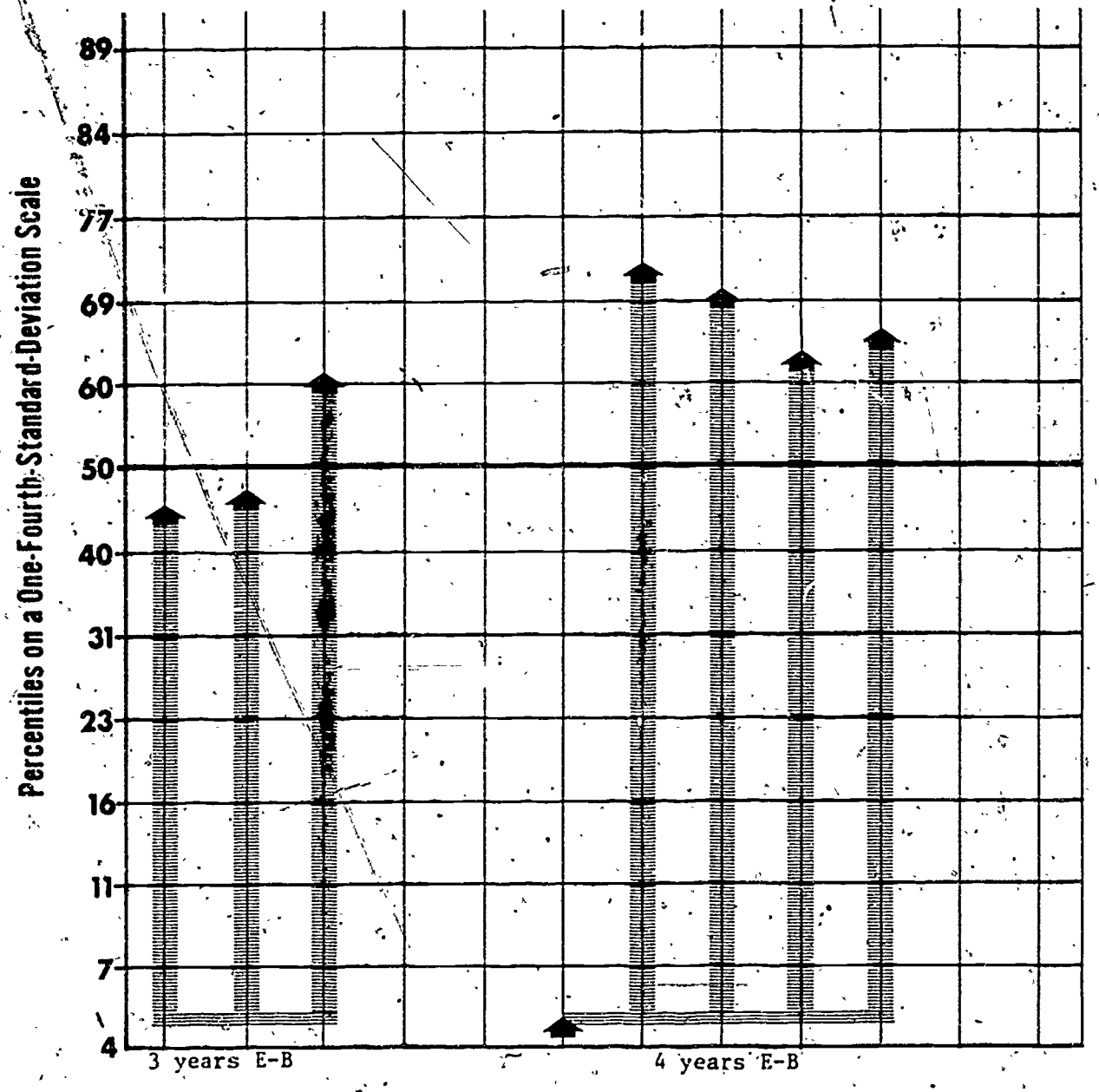
Comparison of Progress for Low-Income E. St. Louis Students with 3 and 4 Years of E-B Follow Through on WRAT Arithmetic (Retentions with Grademates)



PRE	-	-	-	K	-	-	-	-
POST	1	2	3	-	K	1	2	3
N	434	415	388	364	608	619	364	159
%-tile	50	42	50	13	79	68	58	63
S. S.	100	97	100	83	112	107	103	105
S. D.	11.7	8.8	10.8	16.0	14.4	12.3	9.6	8.1
G. E.	1.91	2.78	3.88	.06	1.50	2.34	3.22	4.26

Table 5.8

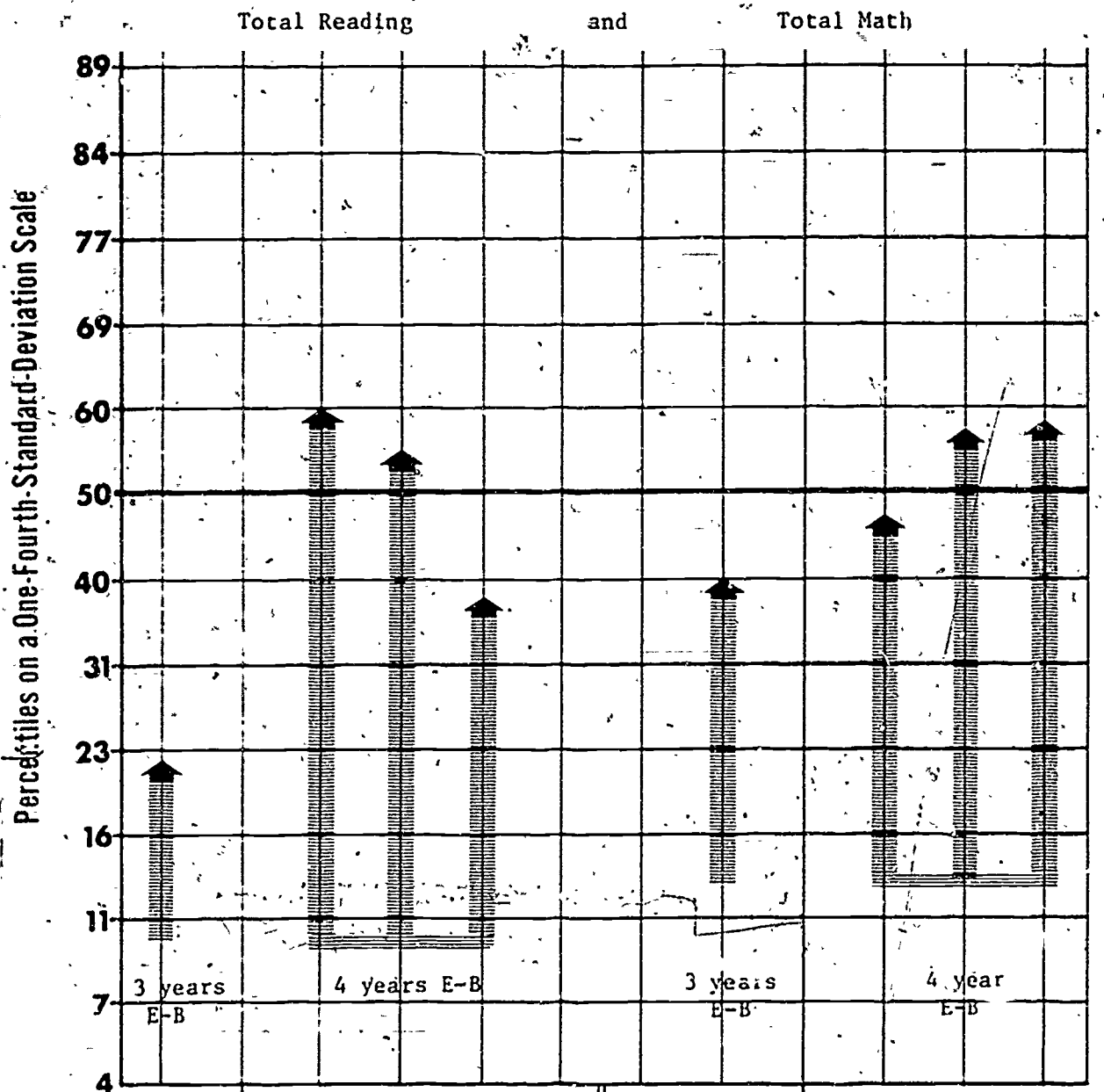
Comparison of Progress for Low-Income E. St. Louis Students with 3 and 4 Years of E-B Follow Through on WRAT Spelling
(Retentions with Grademates)



PRE				K	–	–	–	–
POST	1	2	3	–	K	1	2	3
N	439	420	389	322	632	615	368	160
%-tile	45	47	61	5	73	70	63	66
S. S.	98	99	104	76	109	108	105	106
S. D.	12.5	15.1	19.6	22.2	16.6	15.0	17.8	19.6
G. E.	1.04	2.84	3.66	-.25	1.39	2.36	3.40	4.42

Table 5.9

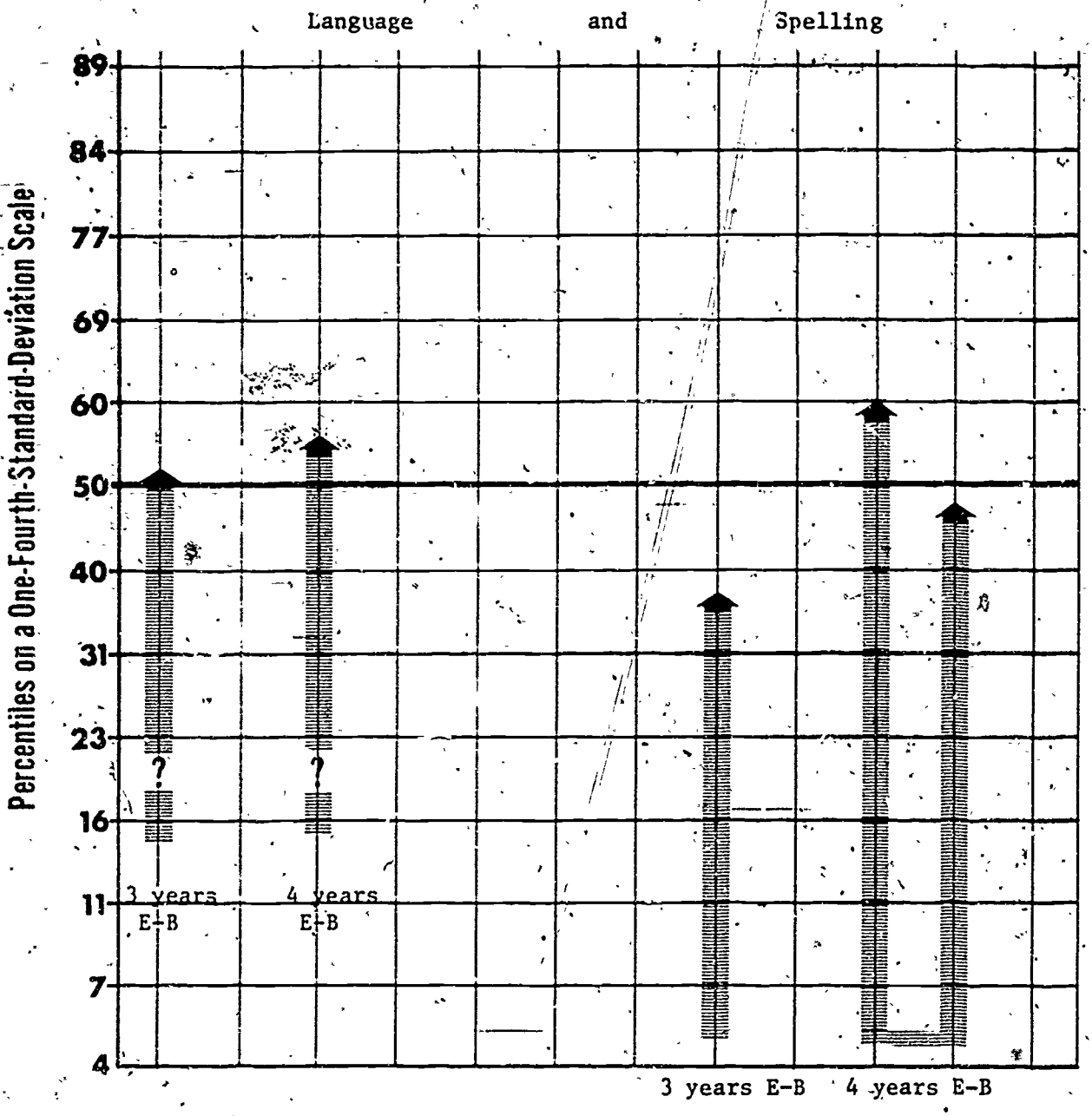
Comparison of Progress for Low-Income E. St. Louis Students with 3 and 4 Years of E-B Follow Through on MAT



PRE	-							
POST	3	1	2	3	3	1	2	3
N	108	507	370	160	106	509	362	159
%-tile	22	60	55	37	40	47	57	58
S. S.	52.1	42.2	53.7	57.4	68.1	40.3	57.1	72.6
S. D.	10.5	10.1	7.3	8.8	11.4	9.6	9.1	9.4
G. E.	2.79	2.09	2.84	3.21	3.69	1.78	2.78	4.06

Table 5.10

Comparison of Progress for Low-Income E. St. Louis Students with 3 and 4 Years of E-B Follow Through on MAT



POST	3	3	3	2	3
N	108	160	105	370	157
%-tile	52	55	37	60	48
S. S.	71.0	71.8	59.6	59.0	63.2
S. D.	12.4	11.7	10.8	11.1	11.5
G. E.	4.40	4.50	3.32	3.25	3.73

In summary, the E. St. Louis data support a strong effect of starting the program in Kindergarten, if the usually high performance of cohort 1 is discounted.

FLIPPIN

Flippin is a small rural White community in northern Arkansas. In 1969 we initiated the program in Flippin for kindergarten, first, and second grades. This gave us another chance to compare the effects of 2, 3, and 4 years of program. As can be noted from the average beginning K baselines in Table 5.11 and 5.12, the Flippin children enter with more skills than found for our typical disadvantaged children. They also progress much further.

Tables 5.11 and 5.12 show the comparisons on WRAT Reading, Arithmetic, and Spelling. Note that we have changed the data being displayed from that shown for Cherokee and E. St. Louis. For the first two within-site studies, we wanted to get across the point that the differences attributable to an early start in the program produced larger effects in the earlier grades than remained by the end of third grade. This same conclusion holds for all the sites being compared. The remainder of the comparisons (except for Headstart) will focus on end of third grade comparisons only and will display the data for each cohort. This allows the reader to see consistencies or trends across cohorts.

For each of the Flippin comparisons, the average 4-year program effects are significantly above the average 2- and 3-year effects. For WRAT Reading and Arithmetic, the significance is beyond the .001 level. For Spelling it is beyond the .01 level. The Reading scores show progressive improvements which are educationally significant. Arithmetic shows a significant gain in going from 2 to 3 years of the program, but not to 4.

The results for Spelling are like those for Arithmetic. The Flippin children are performing well above grade level in all areas.

- - - - - - - - - - -

Tables 5.11 and 5.12

- - - - - - - - - - -

RACINE

Data comparing 3-and 4-year effects from Racine, Wisconsin, are presented in Tables 5.13 and 5.14. The difference for 3-and 4-year effects are each significant. For Reading, the significance is beyond the .05 level, for Arithmetic and Spelling the significance is beyond the .001 level. After the cohort 1 surge which we have seen many times before, there is progressive improvement over cohorts 2 to 4. The magnitude of the difference in effect for K-starting and 1st-starting cohorts is of the order of one-half to three-fourths standard deviation, or .6 to 1.0 grade equivalents.

- - - - - - - - - - -

Tables 5.13 and 5.14

- - - - - - - - - - -

PROVIDENCE

Data comparing 3-and 4-year effects in Providence, Rhode Island are shown in Tables 5.15 and 5.16. The differences in overall effects are each significant beyond the .001 level and amount to at least one-half standard deviation. The children with 4 years of the program are .6 to .8 grade levels ahead of those with only 3 years of the program.

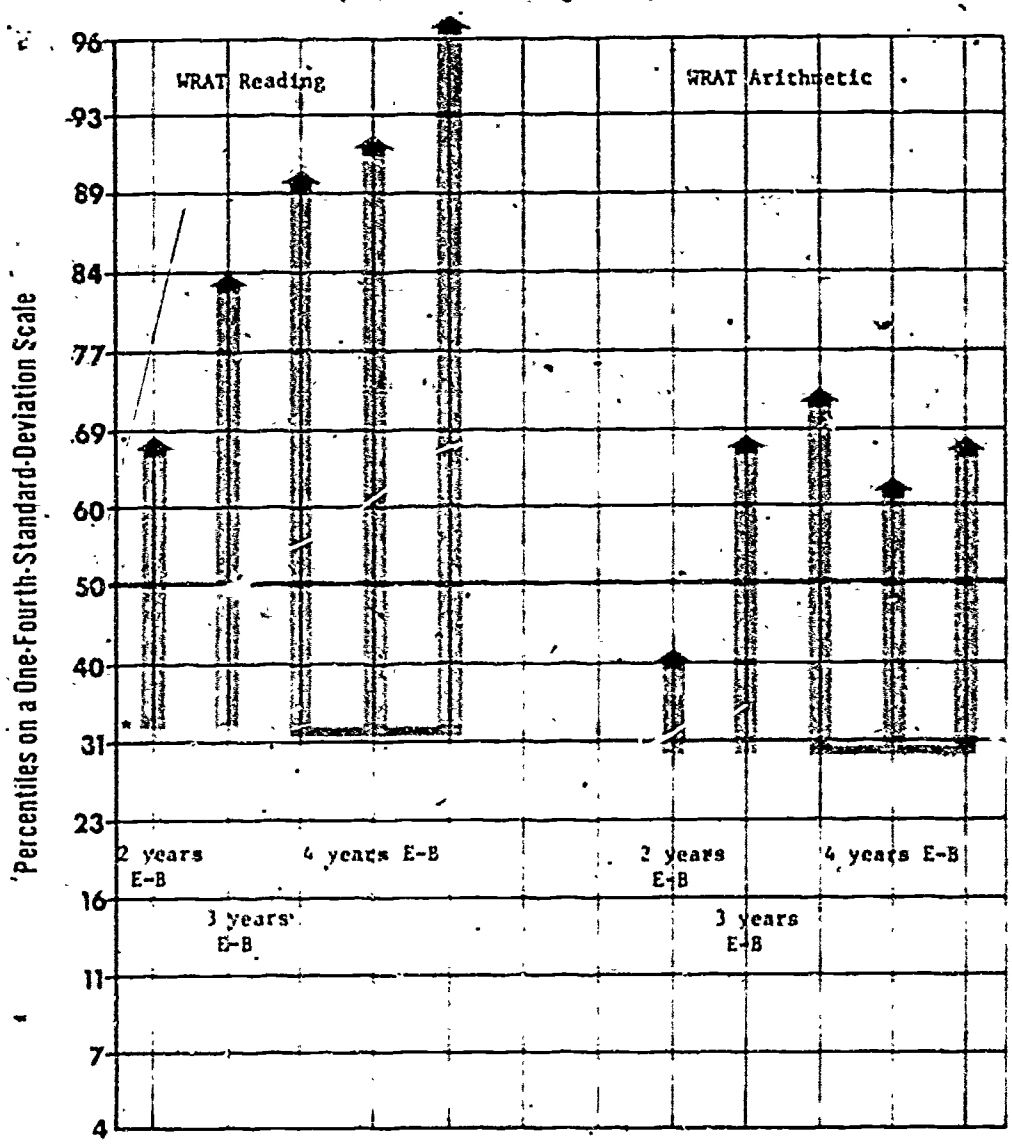
- - - - - - - - - - -

Tables 5.15 and 5.16

- - - - - - - - - - -

Table 5.11

Comparison of Progress at the End of Third Grade for Low-Income Students from Flippin with 2, 3, and 4 Years of E-B Follow Through on WRAT Reading and Arithmetic (Retentions with Agemates)

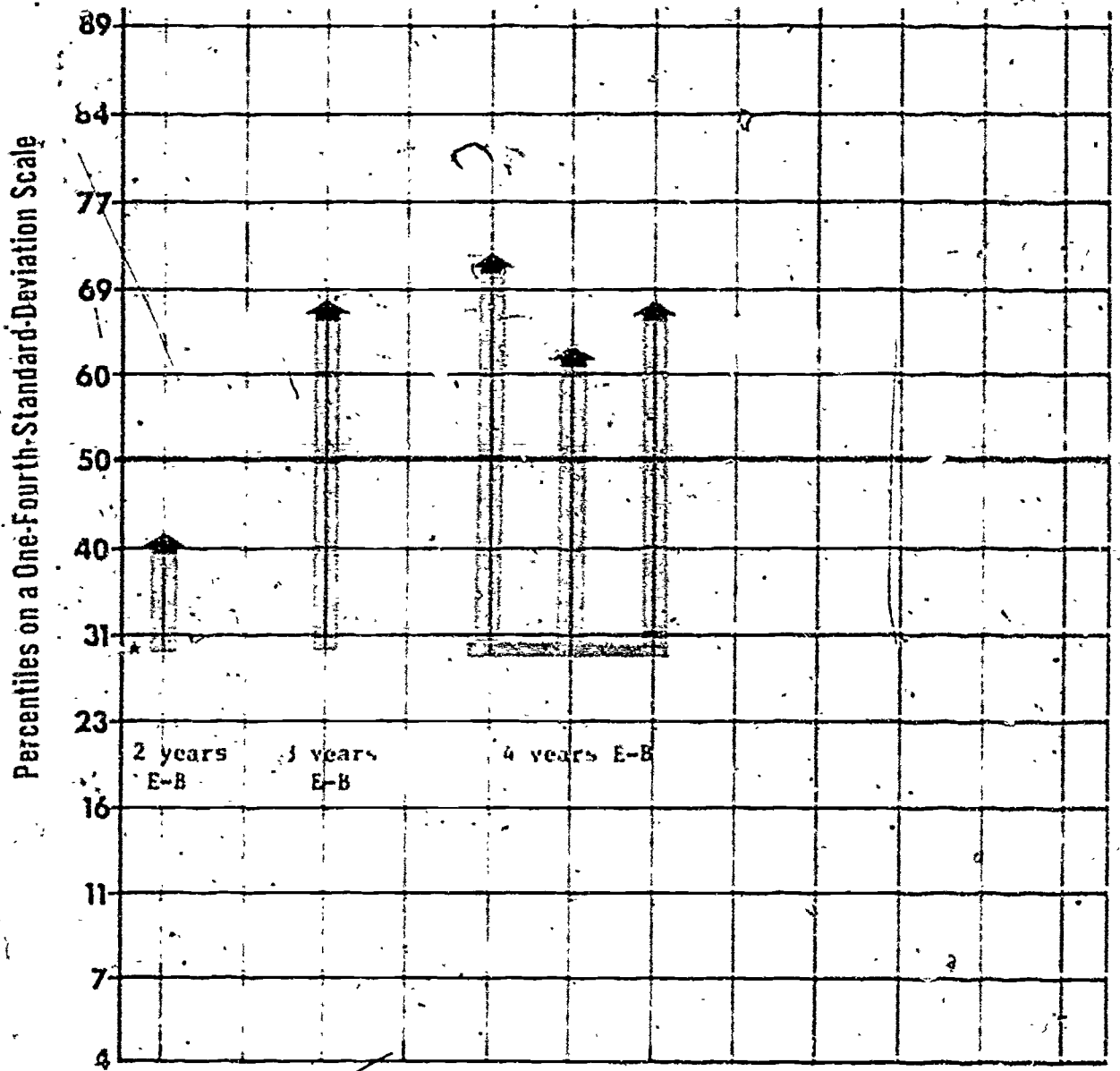


Cohort	X	2	2	3	4	X	2	2	3	4
POST	3	3	3	3	3	3	3	3	3	3
N	20	22	21	18	23	20	22	20	18	23
%-tile	68	84	90	92	99	42	68	73	63	68
S. S.	107	115	119	121	137	97	107	109	105	107
S. D.	19.7	18.3	19.2	16.8	25.5	8.6	6.5	6.0	7.6	7.1
G. E.	4.75	5.25	5.41	6.29	7.82	3.74	4.49	4.49	4.53	4.56

*k-starting baseline.

Table 5.12

Comparison of Progress at the End of Third Grade for Low-Income Students from Flippin with 2, 3, and 4 Years of E-B Follow Through on WRAT Spelling (Retentions with Agemates)

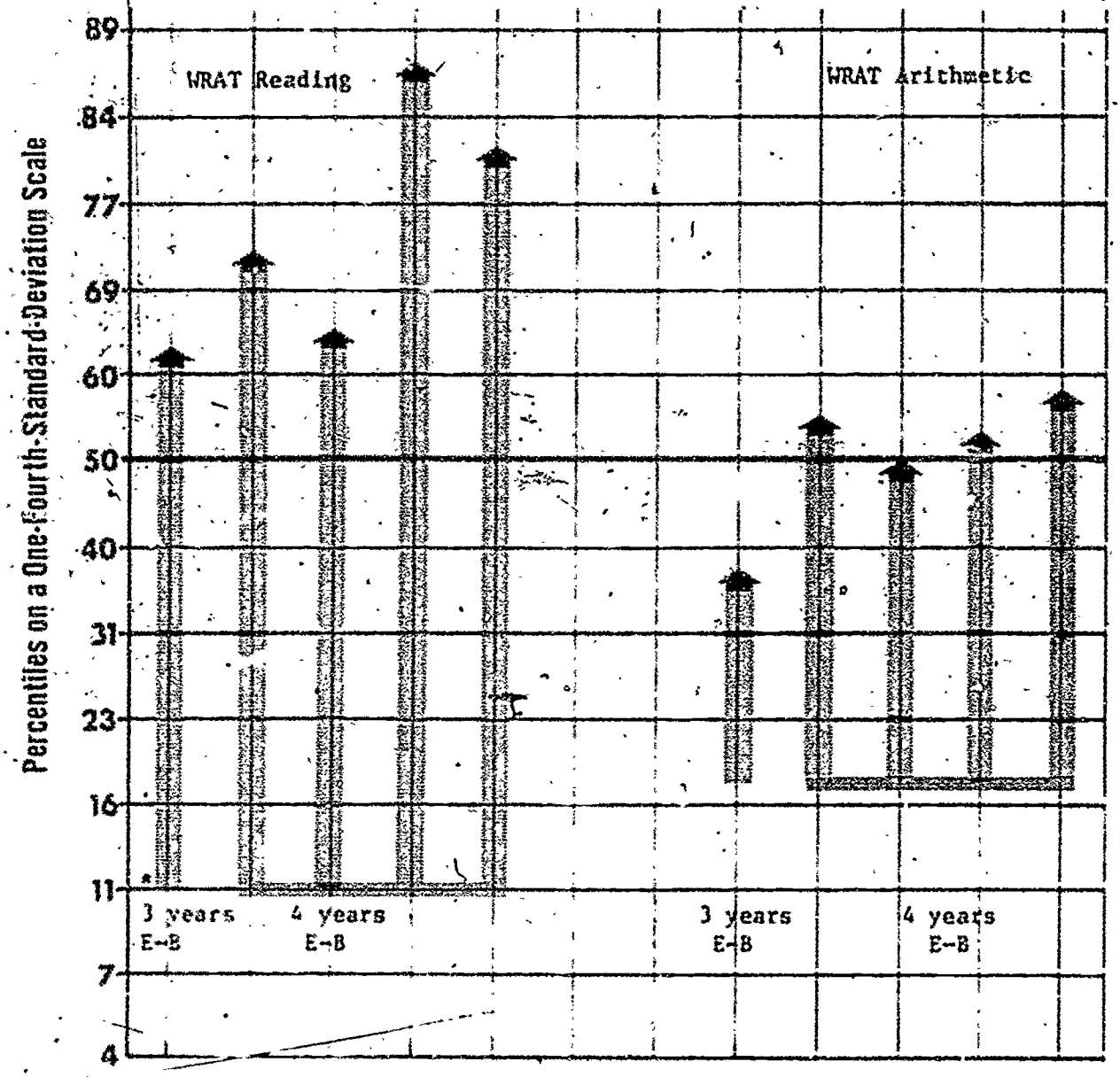


Cohort	X	2	3	4	
POST	3	3	3	3	3
N	29	22	20	18	23
%-tile	42	68	73	84	68
S. S.	9[illegible]	207	209	10[illegible]	10[illegible]
S. D.	8.6	[illegible]	6.6	[illegible]	[illegible]
G. E.	3.74	[illegible]	4.49	4.[illegible]	4.[illegible]

*Ave. = K-starting baseline.

Table 5.13

Comparison of Progress at the End of Third Grade for Low-Income Students from Racine with 3 and 4 Years of E-B Follow Through on WRAT Reading and Arithmetic (Retentions with Agemates)

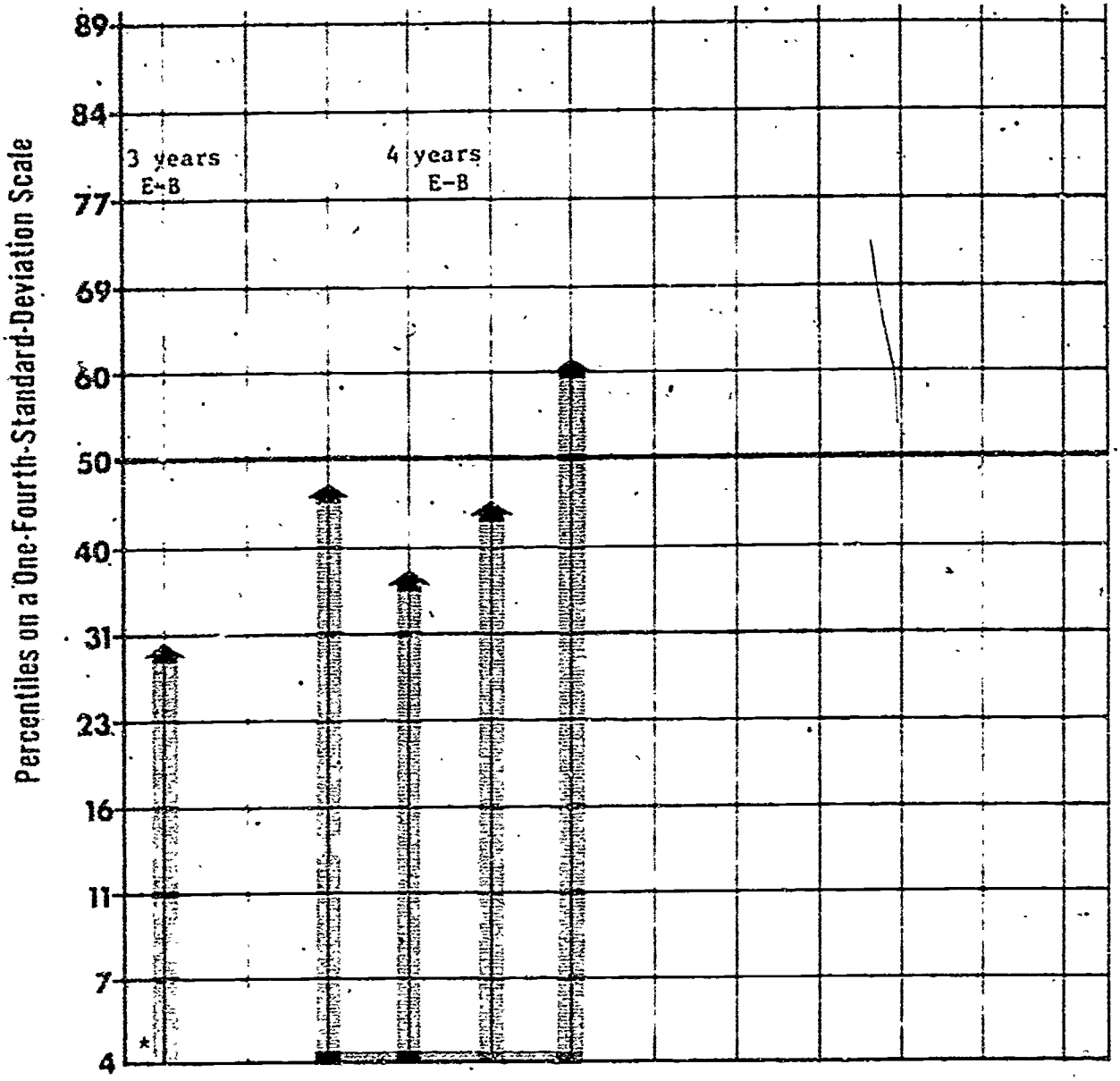


Cohort	1	1	2	3	4	1	1	2	3	4
POST	3	3	3	3	3	3	3	3	3	3
N	43	63	71	56	36	43	63	71	56	37
%-tile	63	73	66	86	81	37	55	50	53	58
S. S.	103	109	106	116	113	95	102	100	101	103
S. D.	16.1	14.7	12.6	20.1	18.6	8.2	10.2	6.9	6.9	8.9
G. E.	4.24	4.68	4.36	5.45	5.13	3.33	3.98	3.78	3.91	4.15

*Average K-starting baseline.

Table 5.14

Comparison of Progress at the End of Third Grade for Low-Income Students from Racine with 3 and 4 Years of E-B Follow Through on WRAT Spelling (Retentions with Agemates)

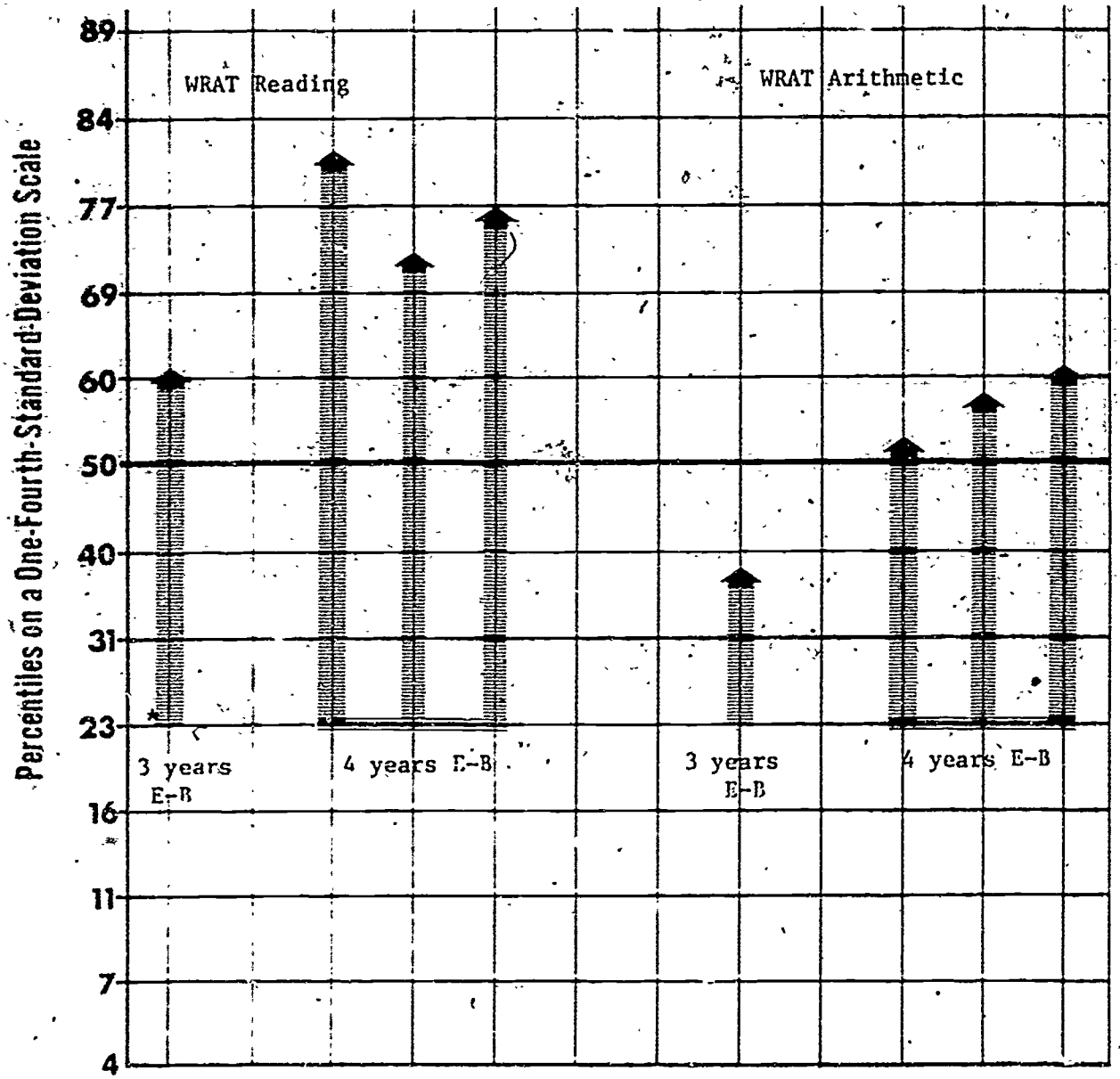


Cohort	1	1	2	3	4
POST	3	4	3	3	3
N	44	63	71	53	37
%-tile	30	47	37	45	61
S. S.	92	99	95	98	104
S. D.	9.4	11.9	9.6	10.9	14.6
G. E.	3.03	3.68	3.34	3.65	4.30

*Average K-starting baseline.

Table 5.15

Comparison of Progress at the End of Third Grade for Low-Income Students from Providence with 3 and 4 Years of E-B Follow Through on WRAT Reading and Arithmetic (Retentions with Agemates)

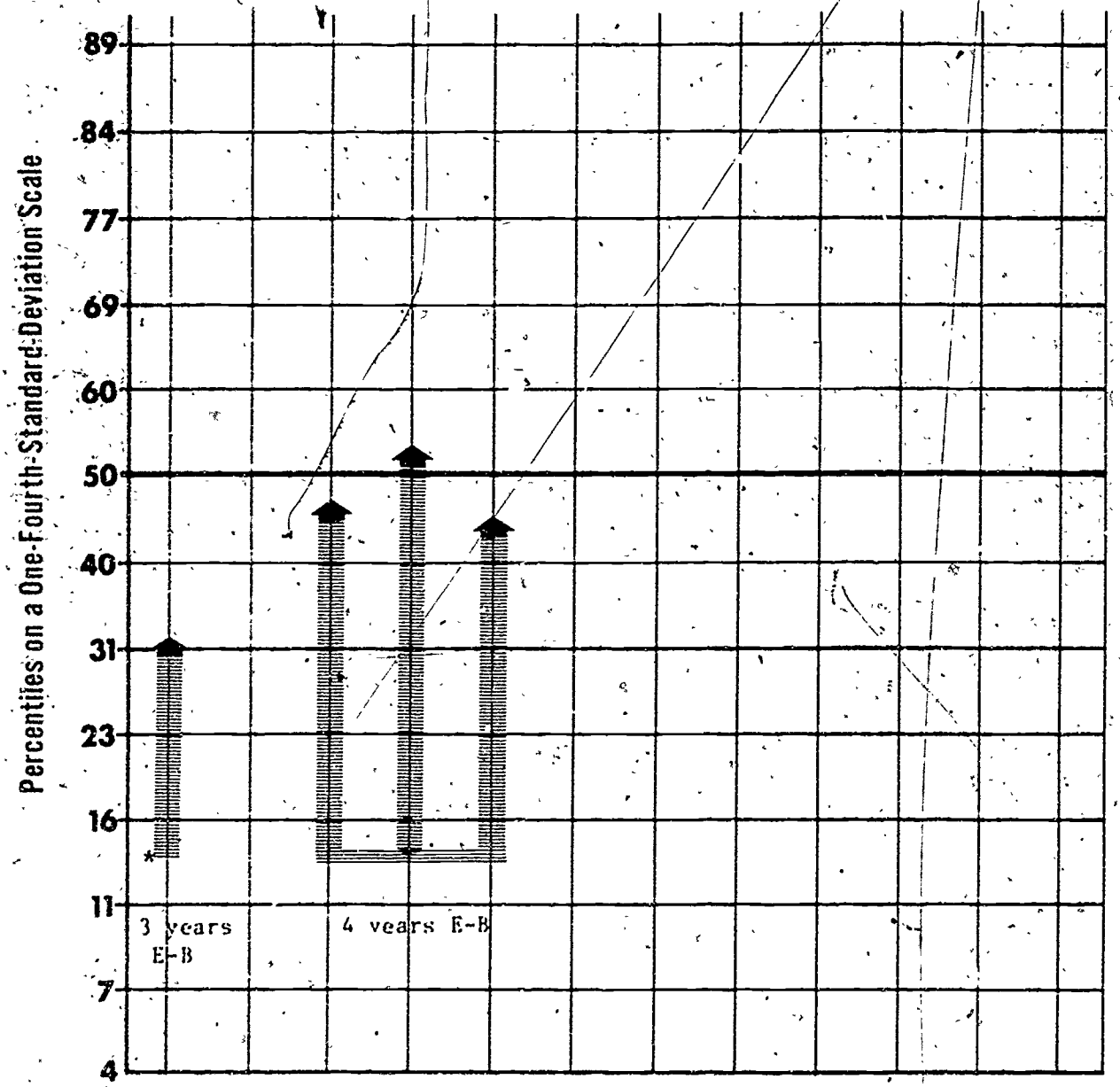


Cohort	2	2	3	4	2	2	3	4
POST	3	3	3	3	3	3	3	3
N	109	78	117	119	109	78	119	116
%-tile	61	81	73	77	37	53	58	61
S.S.	104	113	109	111	95	101	103	104
S.D.	14.8	16.6	16.4	20.2	7.4	8.5	8.4	10.2
G.E.	4.18	5.07	4.65	4.70	3.26	3.86	4.10	3.99

*Average K-starting baseline.

Table 5.16

Comparison of Progress at the End of Third Grade for Low-Income Students from Providence with 3 and 4 Years of E-B Follow Through on WRAT Spelling



Cohort	2	2	3	4
POST	3	3	3	3
N	109	78	104	118
%-tile	32	47	53	45
S. S.	93	99	101	98
S. D.	11.6	15.0	12.0	13.3
G. E.	3.06	3.72	3.85	3.44

*Average K-starting baseline.

ROSEBUD

The Rosebud Sioux Indians are represented in our model by four separate school systems. Three of the jurisdictions are funded through the Rosebud Tribe and are grouped for this presentation. They are the St. Francis Mission School, the Millette County schools at White River and Norris, and the Millette County school at Wood. The fourth jurisdiction is the Todd County Independent School District (presented next).

Cohort 1 includes 1st-starting children at St. Francis. Cohort 2 includes both 1st-starting and K-starting children at St. Francis and in the Millette County schools. Each of the comparisons of the average 3-year and 4-year effects is significant beyond the .001 level. Progressive improvements over K-starting cohorts 2, 3, and 4 are very impressive for both Reading and Arithmetic. The magnitude of the effects are on the order of a full grade level improvement.

- - - - - - - - - - -

Tables 5.17 and 5.18

- - - - - - - - - - -

TODD COUNTY

The final comparisons are for the mostly Sioux children in five Todd County public schools. The overall comparison of 3-year and 4-year effects is significant only for Arithmetic ($p < .001$). Reading shows a progressive improvement over cohorts 2 to 4, and both 3-year and 4-year cohorts perform at an impressive level. The progressive improvements in Arithmetic represent a gain of more than one-half standard deviation. Spelling shows no improvement over cohorts.

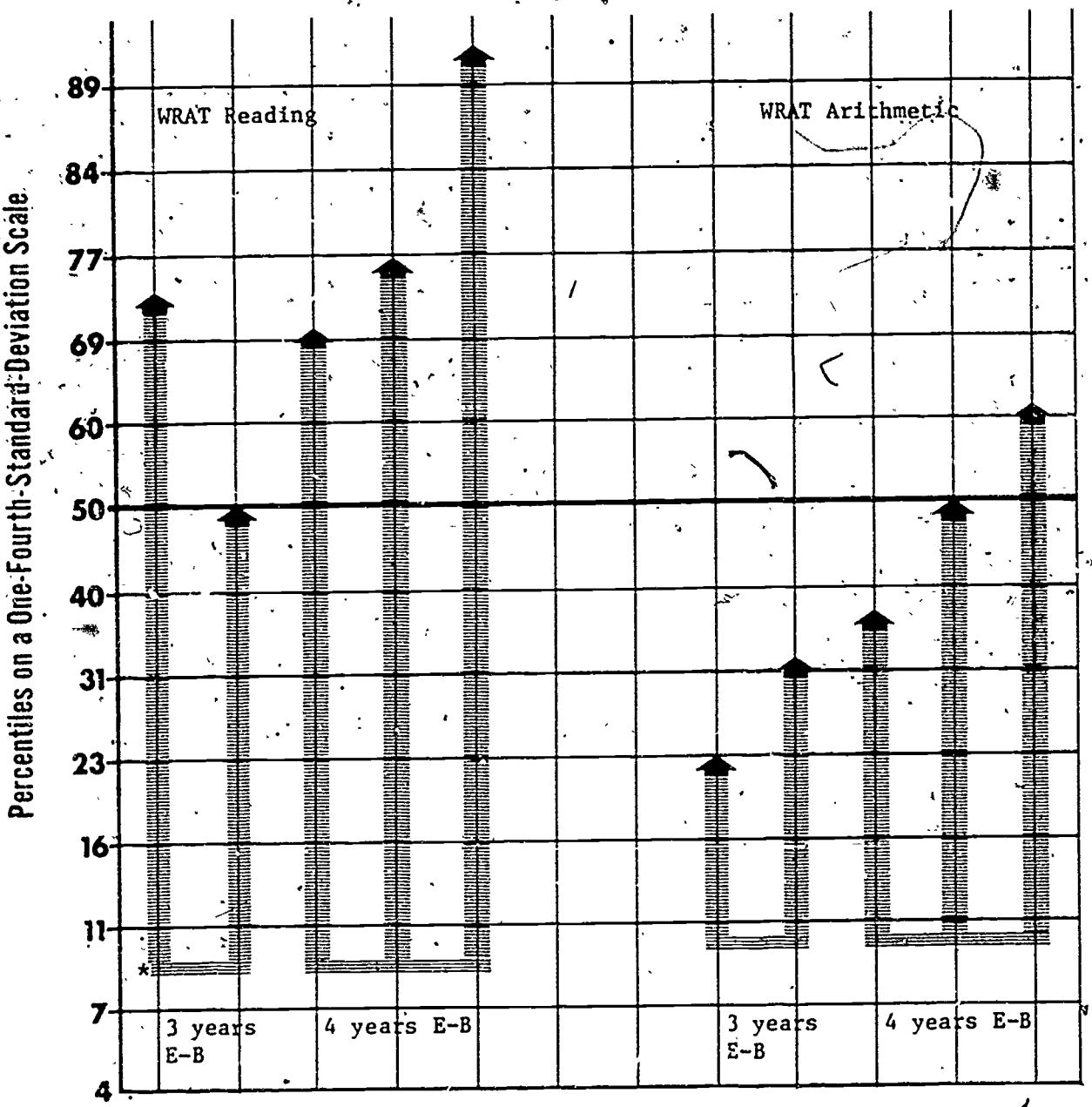
- - - - - - - - - - -

Tables 5.19 and 5.20

- - - - - - - - - - -

Table 5.17

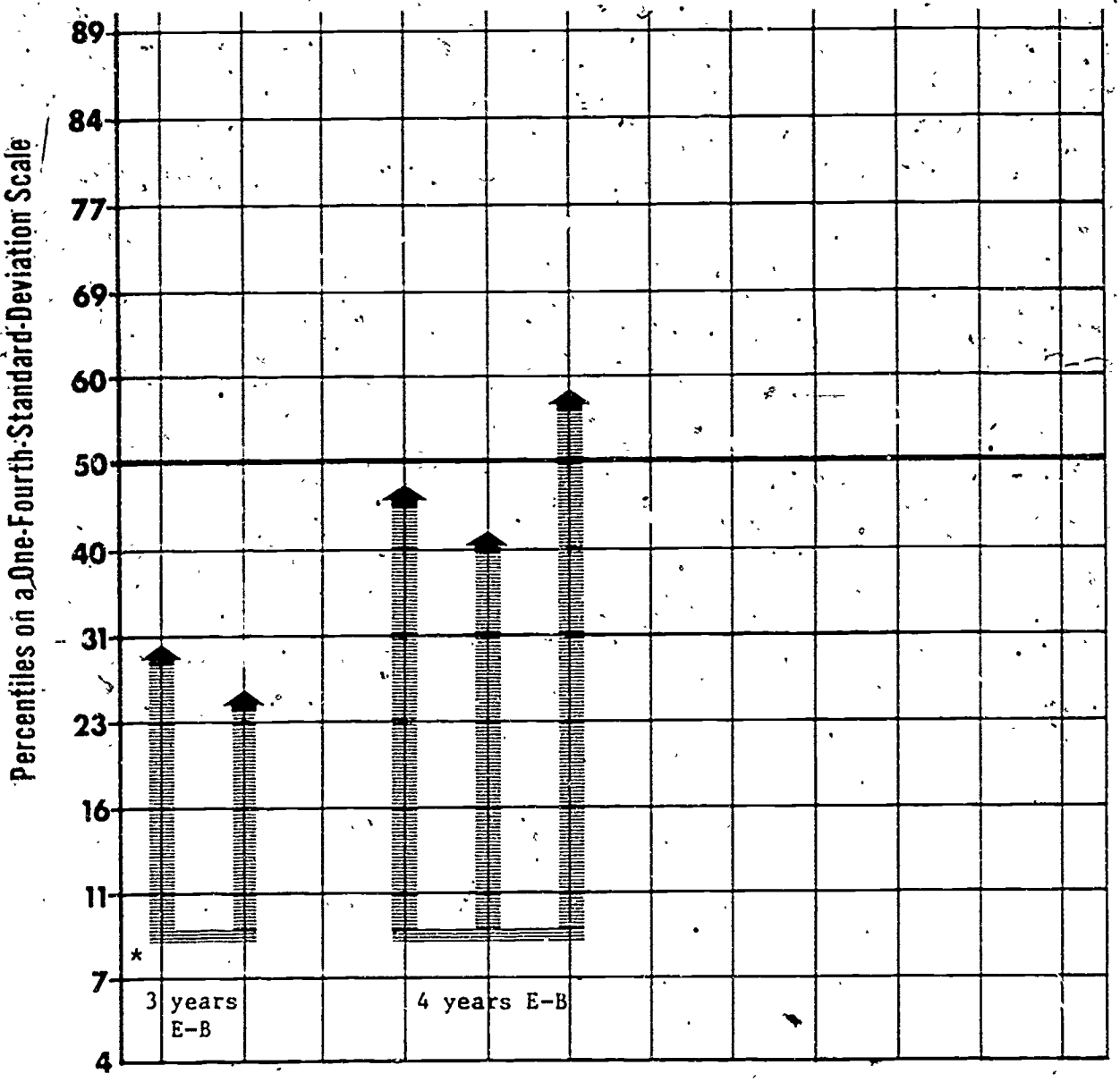
Comparison of Progress at the End of Third Grade for Low-Income Students from Rosebud with 3 and 4 Years of E-B Follow Through on WRAT Reading and Arithmetic (Retentions with Agemates)



Cohort	1	2	2	3	4	1	2	2	3	4
POST	3	3	3	3	3	3	3	3	3	3
N	23	60	32	52	54	23	60	32	52	54
%-tile	73	50	70	77	91	23	32	37	50	61
S. S.	109	100	108	111	120	89	93	95	100	104
S. D.	19.3	11.9	14.0	17.3	21.5	12.2	8.5	8.0	7.8	7.9
G. E.	4.69	4.01	4.59	5.02	5.76	2.76	3.30	3.36	3.97	4.14

Table 5.18

Comparison of Progress at the End of Third Grade for Low-Income Students from Rosebud with 3 and 4 Years of E-B Follow Through on WRAT Spelling (Retentions with Agemates)

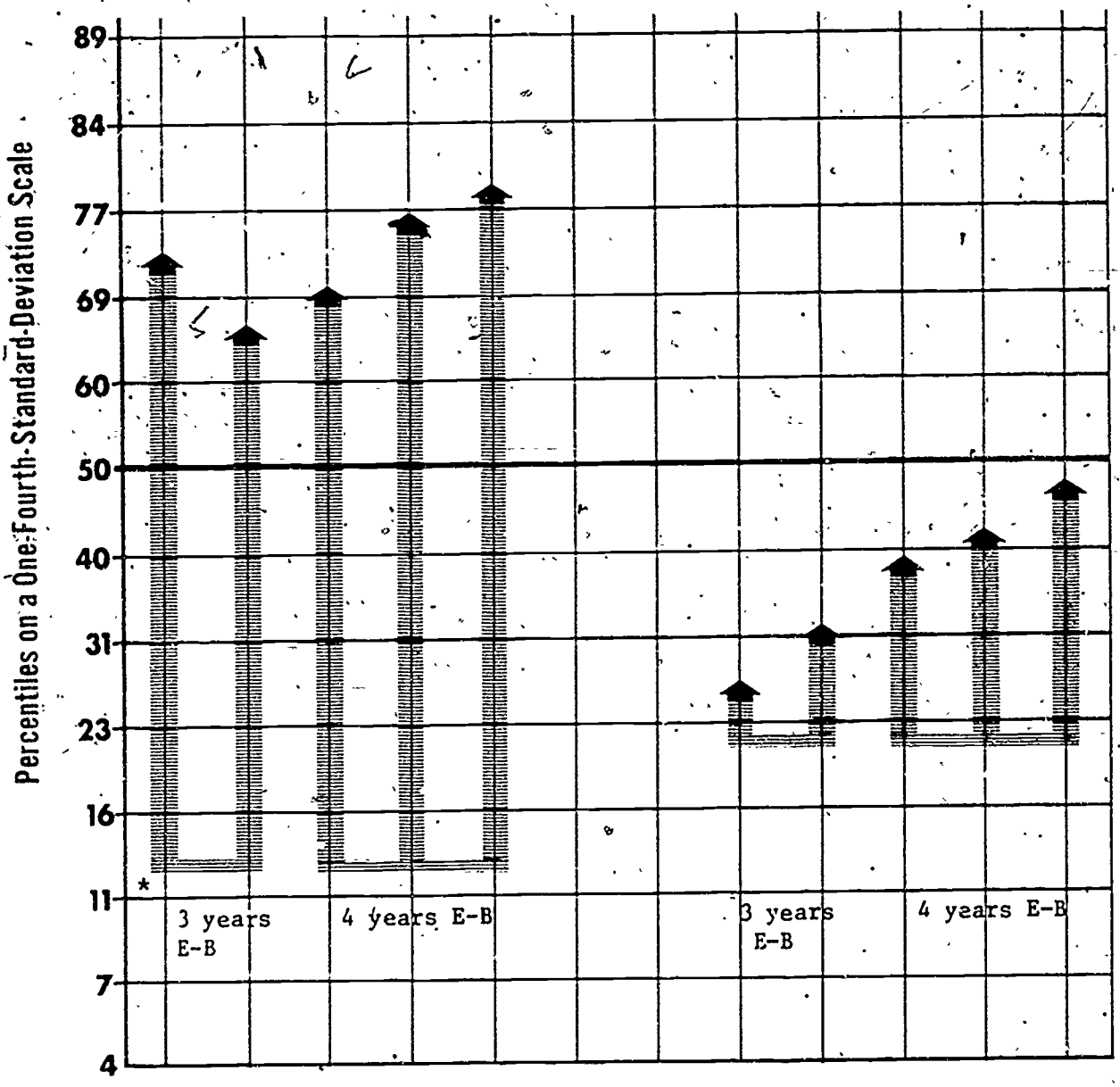


Cohort	1	2	2	3	4
POST	3	3	3	3	3
N	23	60	32	52	54
%-tile	30	25	47	42	58
S. S.	92	90	99	97	103
S. D.	13.8	10.2	11.8	12.5	14.6
G. E.	3.04	2.91	3.70	3.67	4.02

*Average K-starting baseline.

Table 5.19

Comparison of Progress at the End of Third Grade for Low-Income Students from Todd County with 3 and 4 Years of E-B Follow Through on WRAT Reading and Arithmetic (Retentions with Agemates)

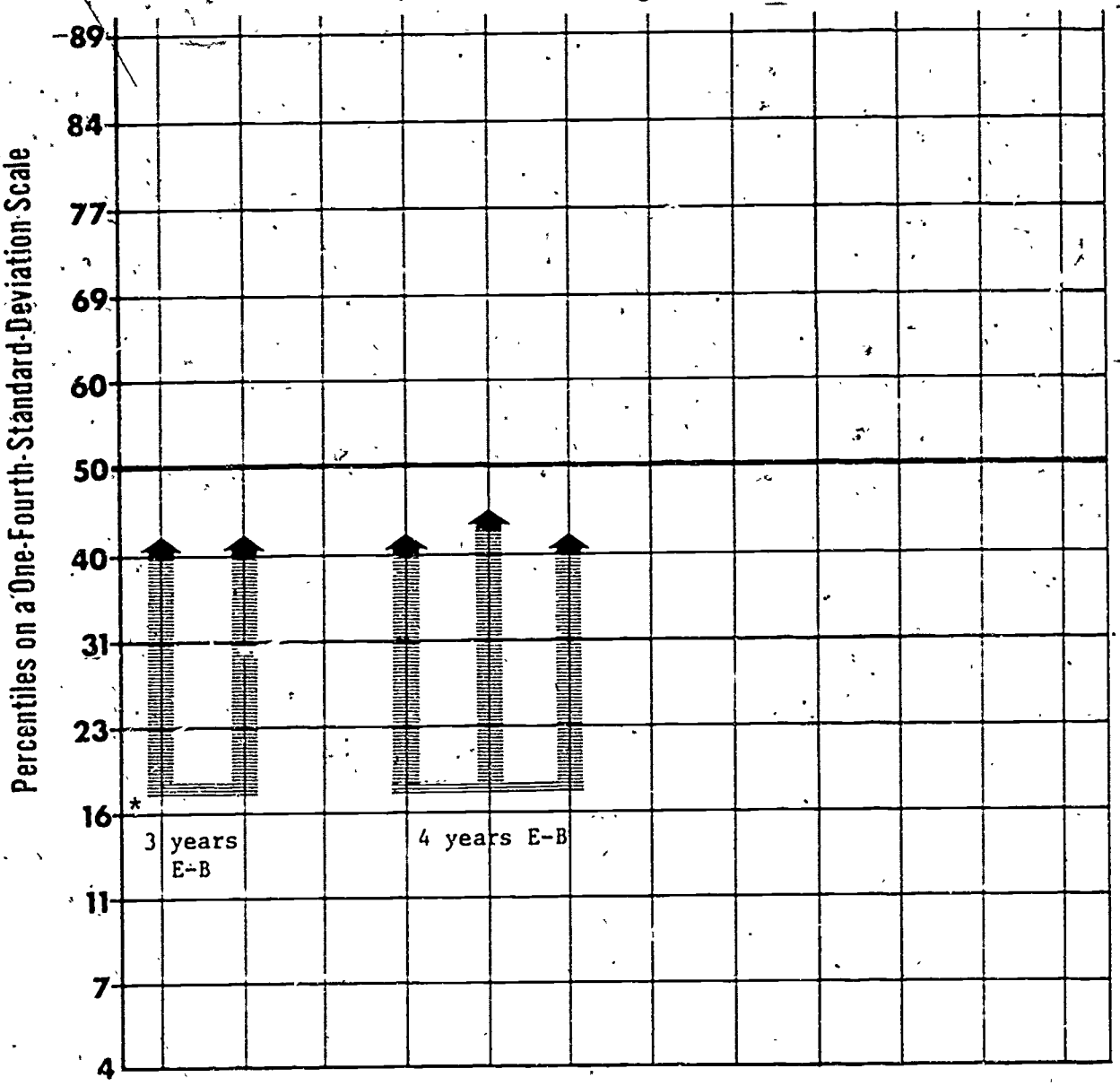


Cohort	1	2	2	3	4	1	2	2	3	4
POST	3	3	3	3	3	3	3	3	3	3
N	79	112	84	92	112	79	113	84	95	110
%-tile	73	66	70	77	79	27	32	39	42	47
S. S.	109	106	108	111	112	91	93	96	97	99
S. D.	16.3	15.6	17.3	19.3	20.0	7.5	8.1	9.6	7.6	8.3
G. E.	4.83	4.52	4.76	5.15	5.12	3.11	3.28	3.47	3.70	3.81

*Average K-starting baseline.

Table 5.20

Comparison of Progress of the End of Third Grade for Low-Income Students from Todd County with 3 and 4 Years of E-B Follow Through on WRAT Spelling (Retentions with Agemates)



Cohort	1	2	2	3	4
POST	3	3	3	3	3
N	79	113	84	95	110
%-tile	42	42	42	45	42
S. S.	97	97	97	98	97
S. D.	11.6	13.6	13.8	11.8	11.8
G. E.	3.63	3.65	3.57	3.75	3.65

*Average K-starting baseline.

Summary on Within-Site Comparisons

These within-site comparisons based on different durations of program implementation demonstrate an advantage to 4 years of program over 3 years of program on the order of one-half standard deviation, or .6 to .8 grade levels. Larger differences are present in some cases and insignificant differences occur for a few comparisons. The overall findings are consistent with the previous comparison made between K-starting and 1st-starting sites, and support the conclusion that the differences reported earlier cannot be attributed to regional differences in students.

PLANNED VARIATION HEAD START COMPARISONS

We now wish to examine how the above findings within Follow Through compare to those combining Planned Variation Head Start with Follow Through. Planned Variation Head Start was operated through Community Action Agencies rather than the public schools, although the programs were often housed in public schools.

In the years 1969, 1970, and 1971, sponsored Head Start programs were initiated in three E-B Follow Through Communities. The communities were Tupelo, Mississippi; E. Las Vegas, New Mexico; and E. St. Louis, Illinois. All programs began with five-year olds. In the second year of the Planned Variation Head Start in E. St. Louis, the program was moved to four-year olds because public school kindergartens were initiated. The program in Tupelo covered 3 entry groups, while those in E. Las Vegas and E. St. Louis covered only 2. The comparisons we will present will look primarily at performance at the end of third grade. However, school entry performances, with and without Head Start can also be examined, as well as the progress by grade level.

Tupelo

Tables 5.21 to 5.25 present the comparisons for Tupelo. On entry to first grade there is a significant superiority of the Head Start students over the non-Head Start students. At the end of third grade, the children who had Head Start are ahead on all eight measures, and six of the differences are significant. The levels of significance are as follows:

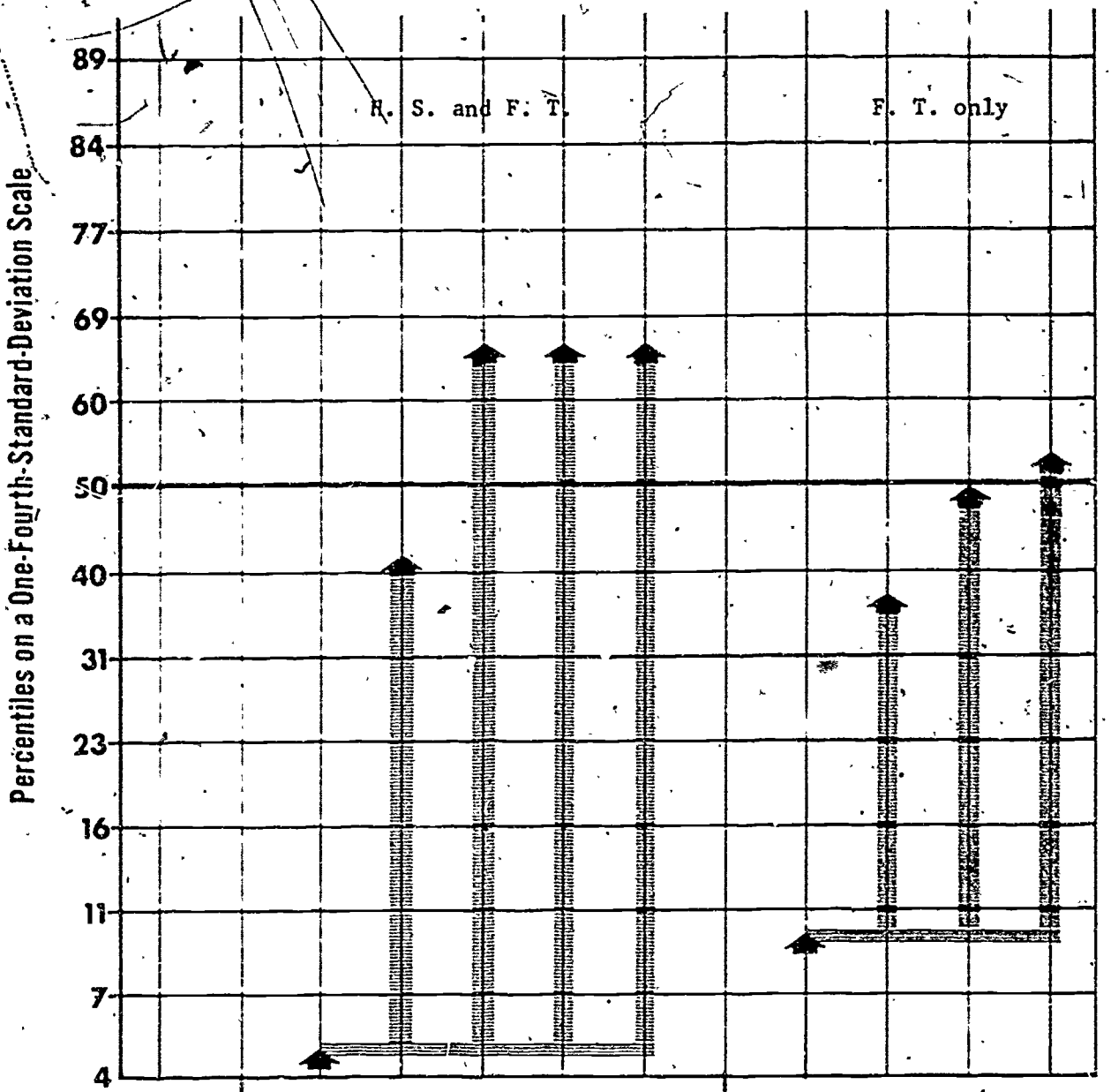
WRAT		
	Reading	.01
	Arithmetic	.001
	Spelling	.001
MAT		
	Total Reading	.001
	Total Math	.001
	Language	N. S.
	Spelling	.001
SIT IQ		N. S.

The magnitude of the significant differences are from a quarter to a half standard deviation, or from .4 to .6 grade levels.

Tables 5.21 to 5.25

Table 5.21

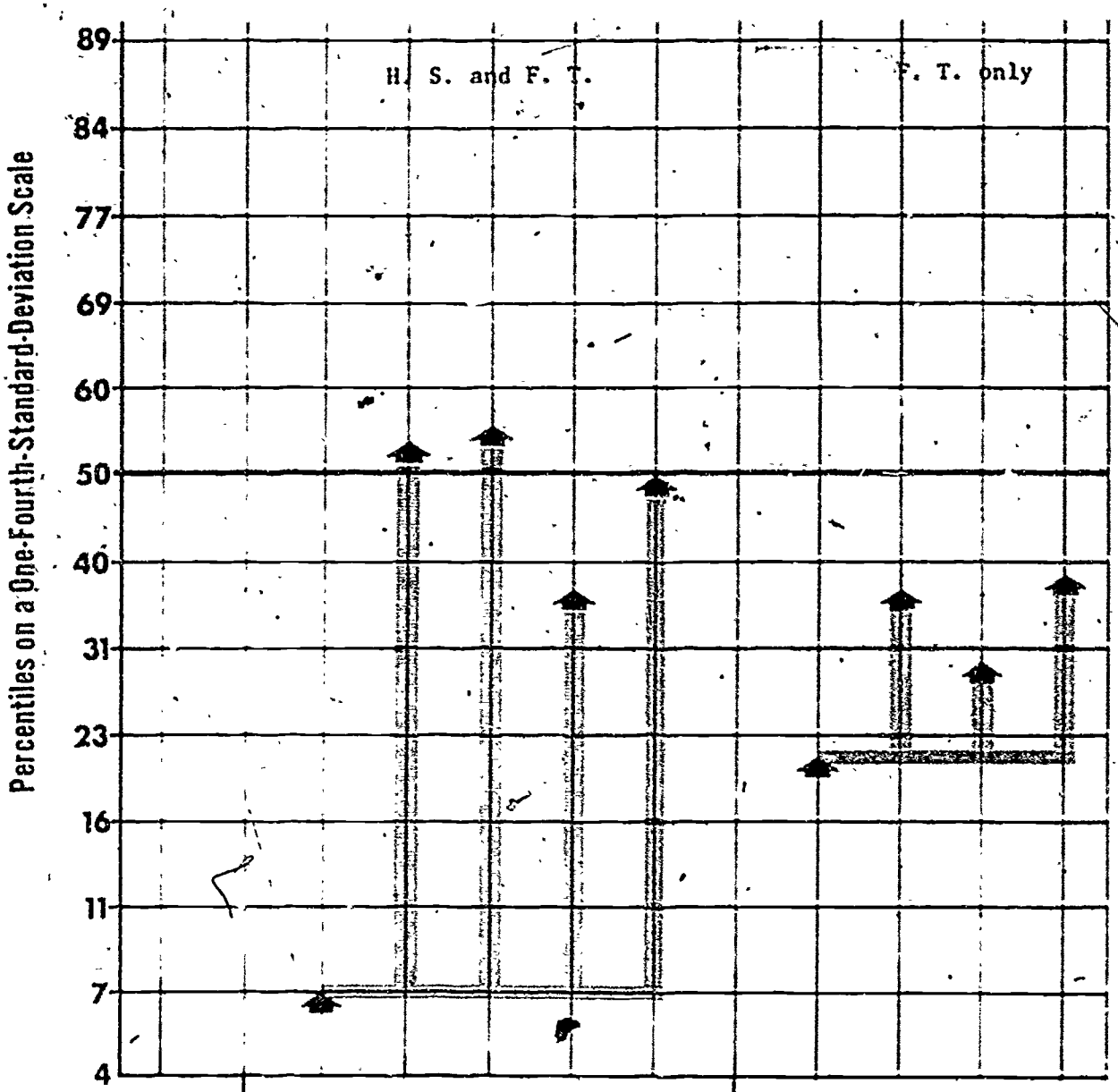
Tupelo Head Start and Follow Through Comparisons on WRAT Reading for Low-Income Children, Retentions with Grademates



PRE	HS	–	–	–	–	1	–		–
POST	–	HS	1	2	3	–	1	2	3
N	65	167	167	84	110	139	364	296	214
%-tile	5	42	66	66	66	10	37	50	53
S. S.	76	97	106	106	106	81	95	100	101
S. D.	13.2	17.9	20.2	18.5	18.2	13.0	16.3	17.8	17.3
G. E.	-.16	.83	2.20	3.40	4.42	.35	1.59	3.04	3.91

Table 5.22

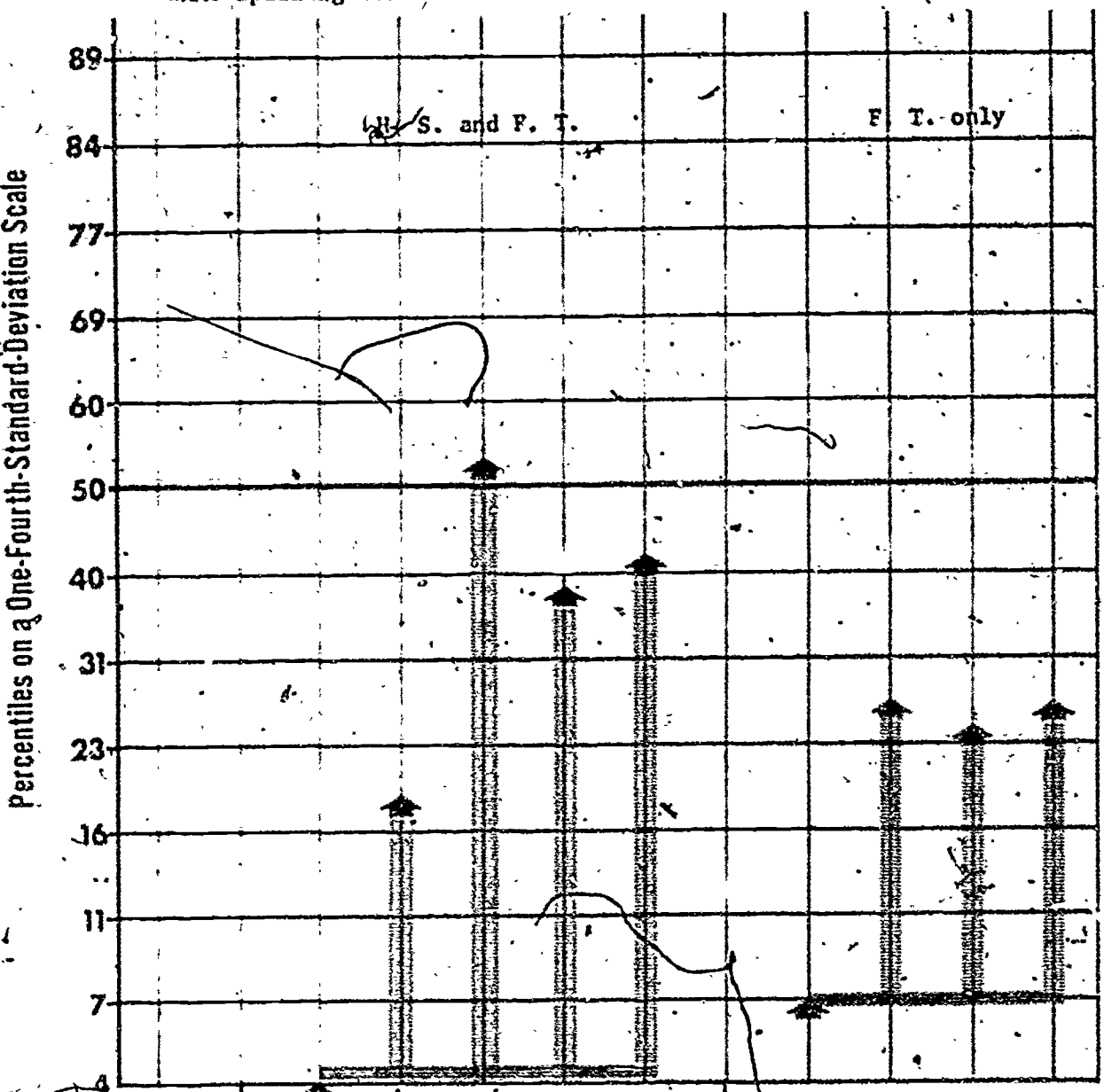
Tupelo Head Start and Follow Through Comparisons on
WRAT Arithmetic for Low-Income Children, Retentions with Grademates



PRE	HS	-	-	-	-	1	-	-	-
POST	-	HS	1	2	3	-	1	2	3
N	73	170	168	84	110	141	363	296	214
%-tile	7	53	55	37	50	21	57	30	39
S. S.	78	101	102	95	100	88	95	92	96
S. D.	15.4	14.3	10.9	7.0	8.2	12.2	11.9	8.3	8.6
G. E.	-.14	1.04	1.90	2.70	3.90	.65	1.54	2.38	3.49

Table 5.23

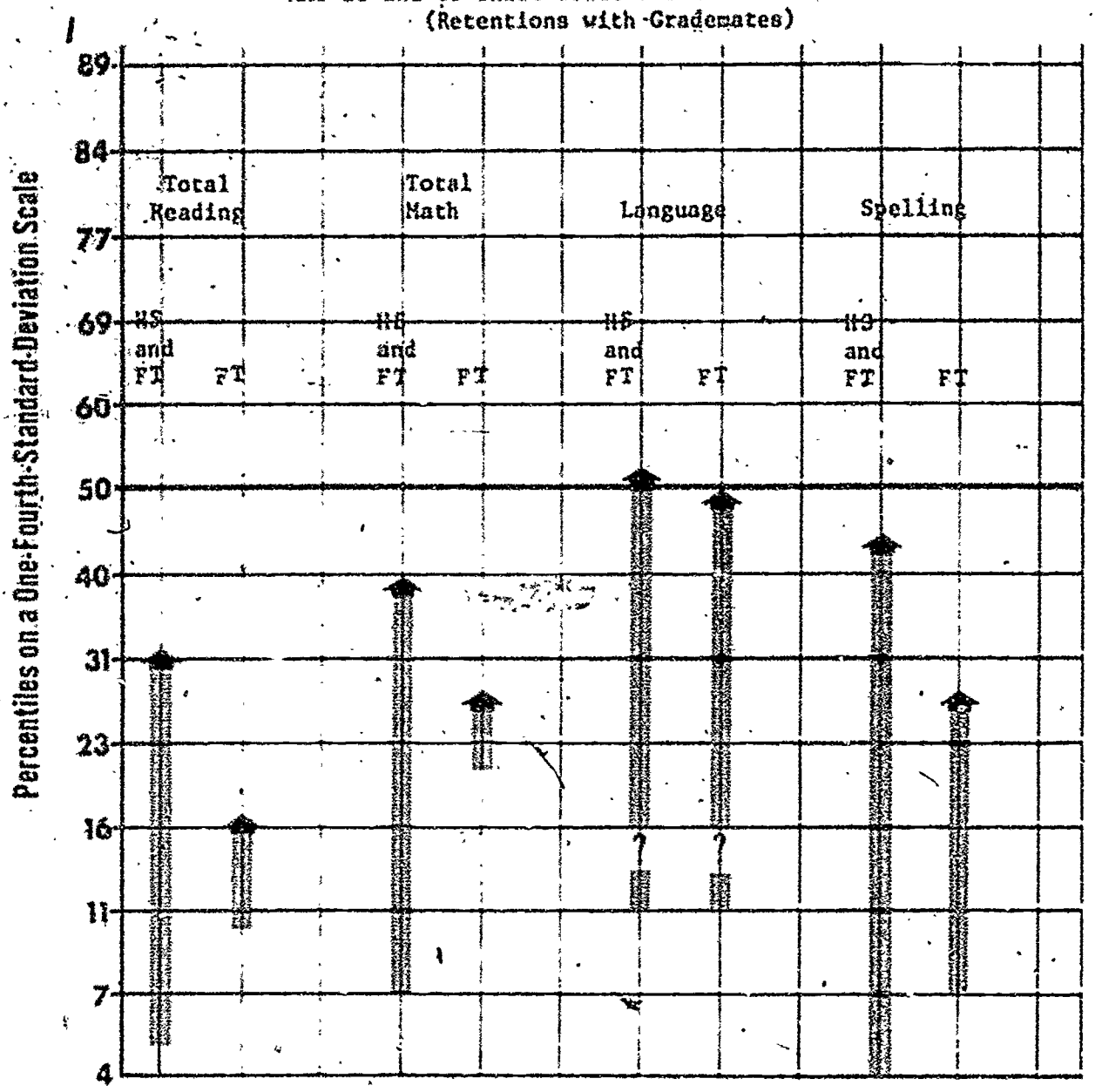
Tupelo Head Start and Follow Through Comparisons on
WRAT Spelling for Low-Income Children, Retentions with Grademates



PRE	HS	-	-	-	-	1	-	-	-
POST	-	HS	1	2	3	-	1	2	3
N	51	168	168	84	107	132	364	296	207
%-tile	1	19	53	39	42	7	27	25	27
S. S.	63	87	101	96	97	78	91	90	91
S. D.	18.6	19.6	13.9	12.4	12.3	16.1	13.7	11.0	11.4
G. E.	-.74	.49	1.86	2.75	3.53	.22	1.36	2.20	2.98

Table 5.24

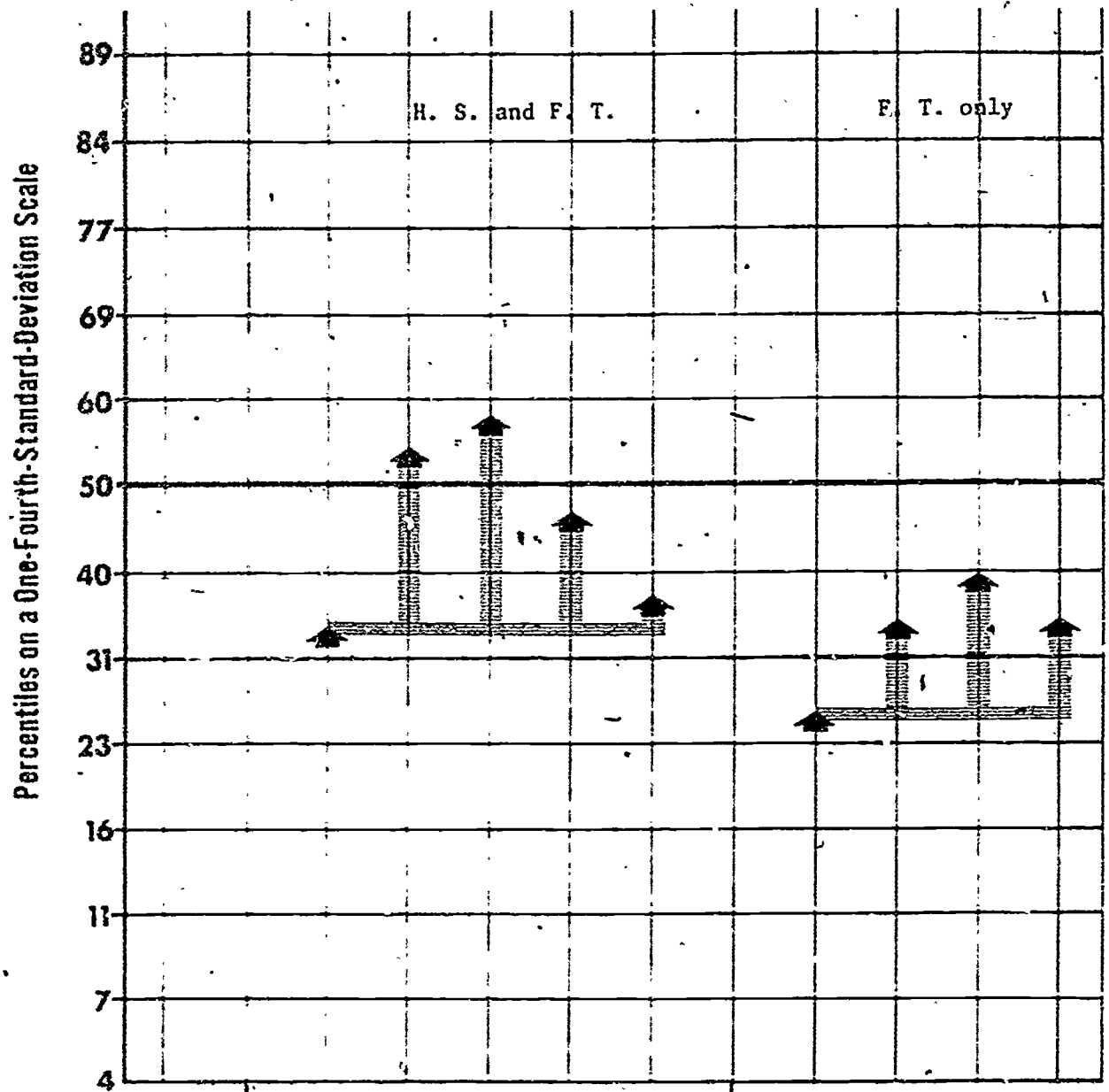
Tupelo Head Start and Follow Through Comparisons on MAT at End of Third Grade for Low-Income Children (Retentions with Grademates)



PRE	HS	1	HS	1	-	-	HS	1
POST	3	3	3	3	3	3	3	3
N	109	80	108	80	110	81	107	74
%-tile	32	17	40	28	52	50	44	28
S. S.	55.4	50.6	68.0	64.1	70.8	69.8	62.6	56.5
S. D.	8.0	7.3	10.0	9.4	12.2	10.7	11.7	10.5
G. E.	3.04	2.59	3.54	3.31	4.48	4.22	3.64	3.02

Table 5.25

Tupelo Head Start and Follow Through Comparisons on SIT IQ's for Low-Income Children, Retentions with Grademates



PRE	HS	-	-	-	-	1	-	-	-
POST	-	HS	1	2	3	-	1	2	3
N	96	170	167	84	110	231	361	293	210
%-tile	34	54	58	47	38	26	35	40	35
S. S.	93.8	101.6	103.1	98.8	95.7	90	94.4	96.5	94.2
S. D.	14.8	15.3	14.8	14.3	16.0	13.0	13.8	13.8	16.4

E. Las Vegas

Tables 5.26 to 5.30 present the Head Start comparisons for E. Las Vegas. Again, on entry to first grade there are clearly significant differences favoring the Head Start children. These differences are reduced in size by the end of third grade, but still all eight measures favor the Head Start children and 5 of the 8 do so significantly. The levels of significance are as follows:

WRAT	
Reading	N. S.
Arithmetic	.05
Spelling	N. S.
MAT	
Total Reading	.05
Total Math	.01
Language	.01
Spelling	N. S.
SIT IQ	.001

The magnitude of the significant differences are on the order of one-quarter standard deviation, or .2 to .4 grade levels.

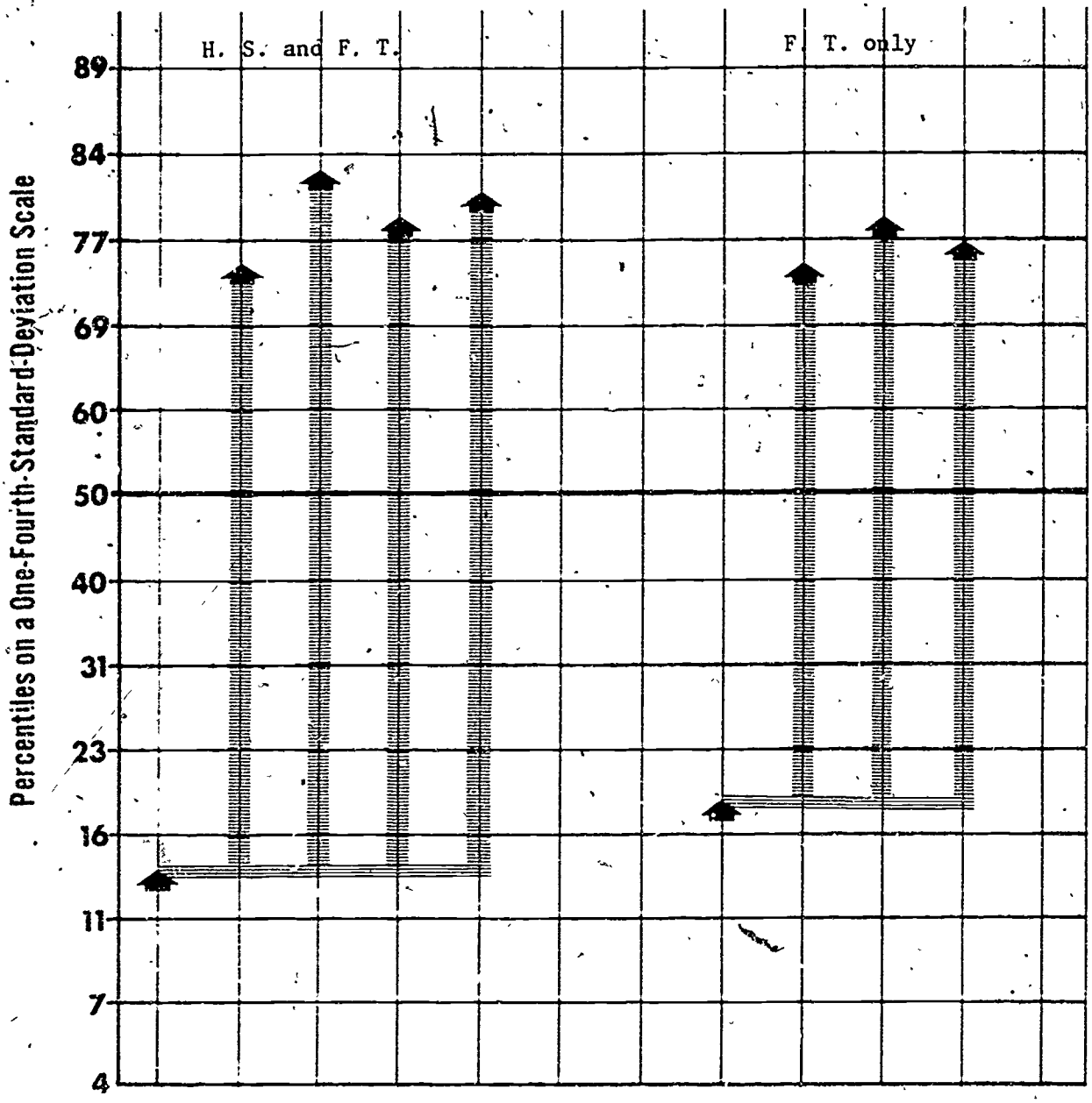
- - - - - - - - - - -

Tables 5.26 to 5.30

- - - - - - - - - - -

Table 5.26

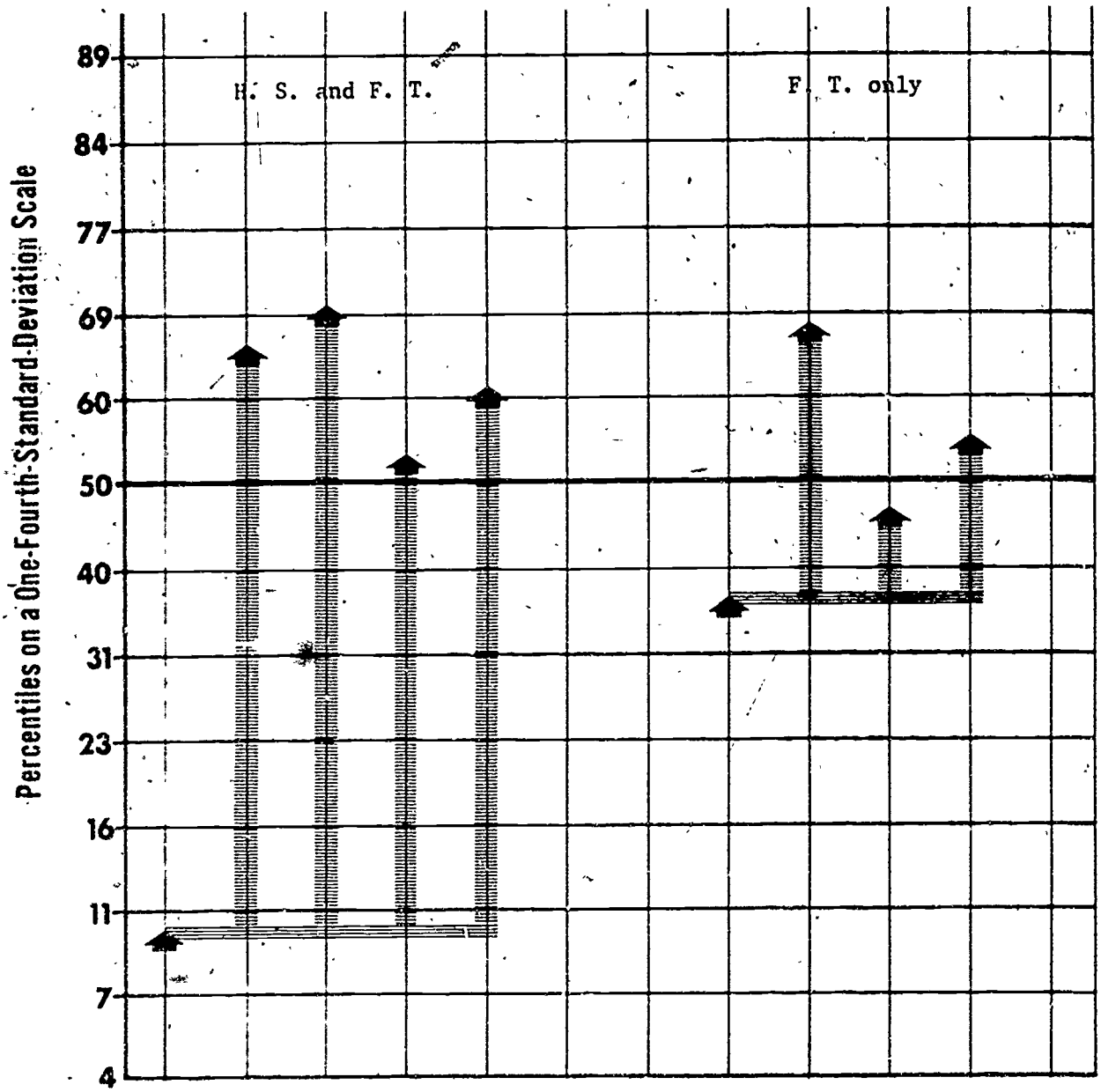
E. Las Vegas Head Start and Follow Through Comparisons on WRAT Reading for Low-Income Students, Retentions with Grademates



PRE	HS	–	–	–	–	1	–	–	–
POST	–	HS	1	2	3	–	1	2	3
N	48	149	149	132	121	259	394	353	246
%-tile	14	75	83	79	81	19	75	79	77
S. S.	84	110	114	112	113	87	110	112	111
S. D.	10.2	20.3	18.3	15.9	17.3	12.5	17.1	16.0	16.4
G. E.	.11	1.37	2.64	3.94	5.06	.53	2.46	3.93	4.93

Table 5.27

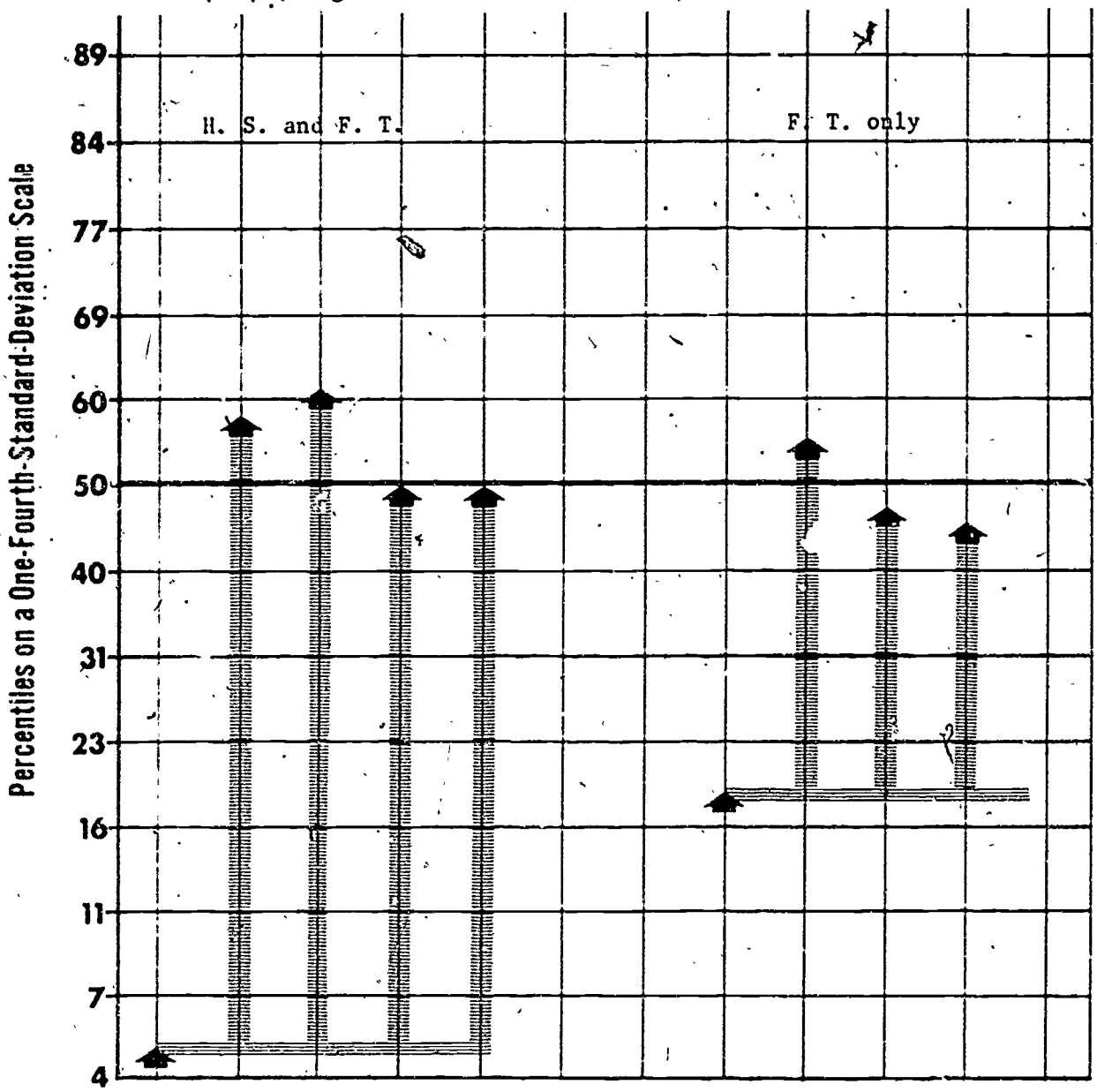
E. Las Vegas Head Start and Follow Through Comparisons on WRAT Arithmetic for Low-Income Students, Retentions with Grademates



PRE	HS	-	-	-	-	1	-	-	-
POST	-	HS	1	2	3	-	1	2	3
N	48	149	149	132	121	269	394	353	246
%-tile	10	66	70	53	61	37	68	47	55
S. S.	81	106	108	101	104	95	107	99	102
S. D.	17.6	13.0	10.4	7.9	7.3	12.2	10.7	7.4	8.4
G. E.	-.04	1.21	2.29	3.04	4.19	.89	2.24	2.88	4.03

Table 5.28

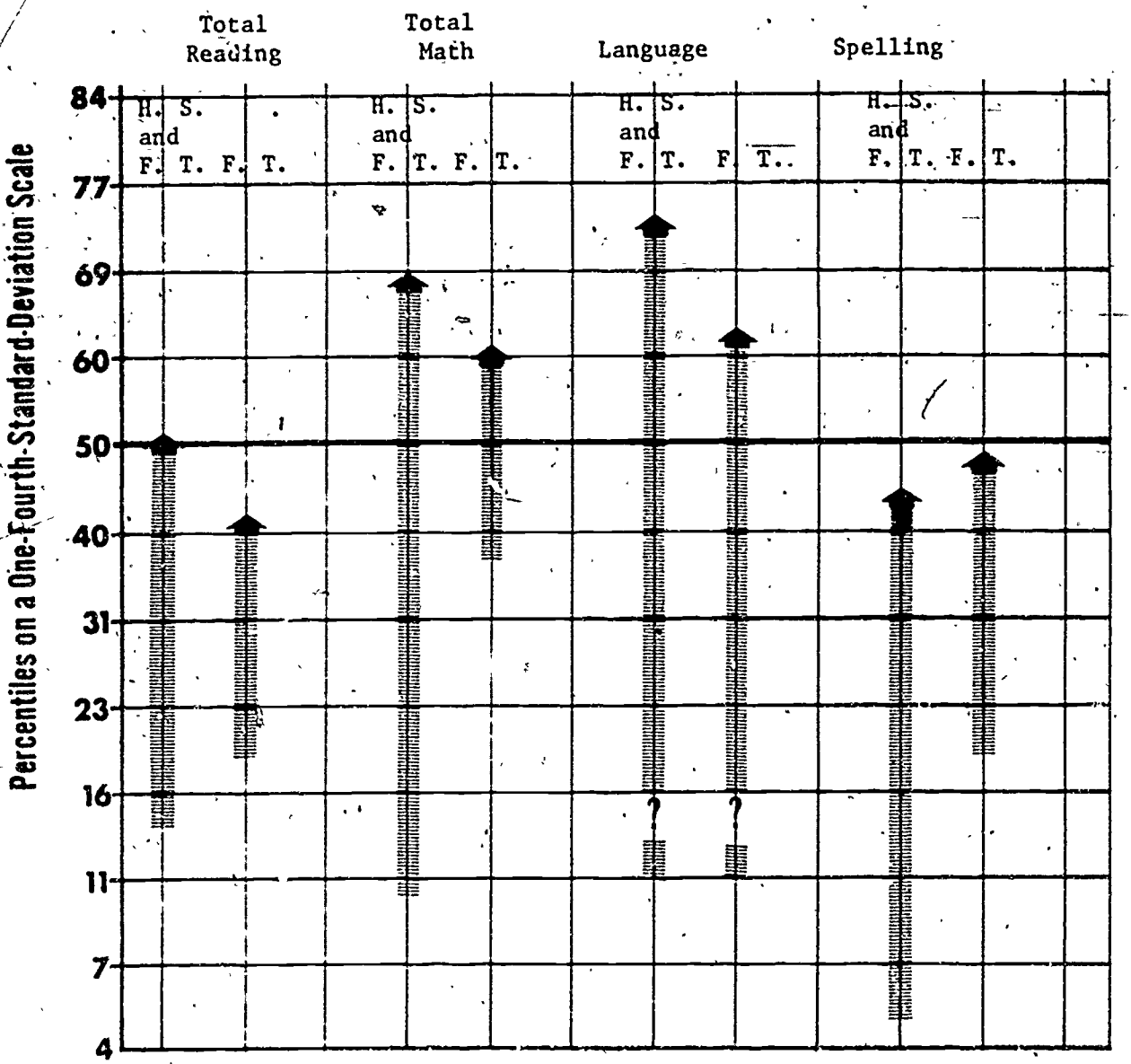
E. Las Vegas Head Start and Follow Through Comparisons on WRAT Spelling for Low-Income Students, Retentions with Grademates



PRE	HS	-	-	-	-	1	-	-	-
POST	-	HS	1	2	3	-	1	2	3
N	44	149	149	132	121	266	396	353	246
%-tile	5	58	61	50	50	19	55	47	45
S. S.	75	103	104	100	100	87	102	99	98
S. D.	19.6	14.6	13.7	13.3	13.4	14.0	11.8	11.6	12.7
G. E.	-.25	1.08	2.07	2.95	3.78	.53	1.95	2.88	3.67

Table 5.29

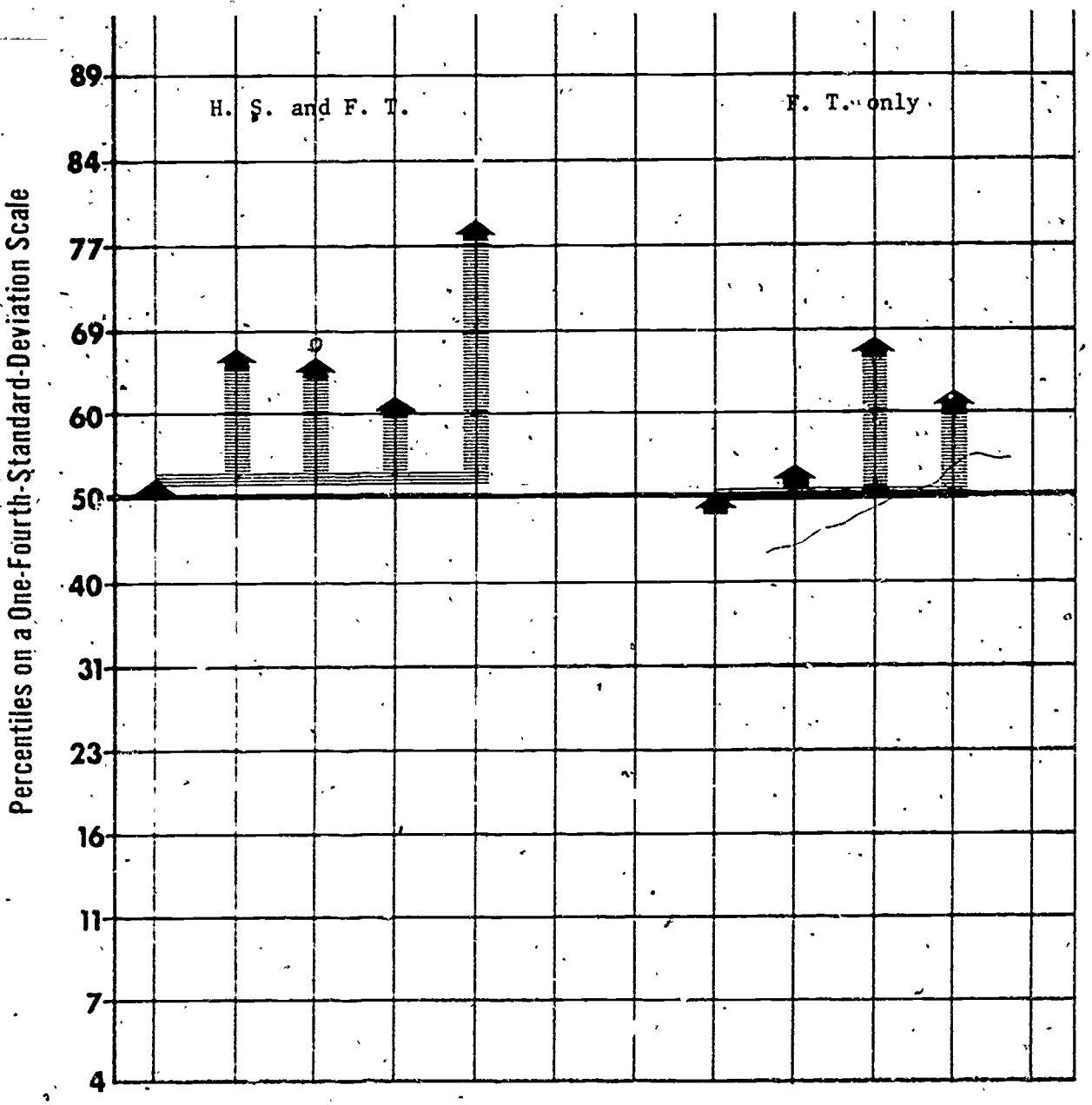
E. Las Vegas Head Start and Follow Through Comparisons on MAT at End of Third Grade for Low-Income Children (Retentions with Grademates)



	Total Reading		Total Math		Language		Spelling	
PRE	HS	1	HS	1	HS	1	HS	1
POST	3	3	3	3	3	3	3	3
N	113	95	121	97	121	97	122	96
%-tile	51	42	69	61	74	63	45	49
S. S.	61.4	59.1	76.5	73.4	78.2	74.6	62.4	63.6
S. D.	8.5	8.8	8.9	8.2	10.5	8.8	10.1	11.8
G. E.	3.61	3.38	4.45	4.12	5.38	4.76	3.61	3.78

Table 5.30

E. Las Vegas Head Start and Follow Through Comparisons on SIT IQ's for Low-Income Children, Retentions with Grademates



PRE	HS	-	-	-	-	1	-	-	-
POST	-	HS	1	2	3	-	1	2	3
N	61	150	84	69	121	267	384	262	249
%-tile	52	67	66	62	79	50	53	68	63
S. S.	100.8	106.3	106.1	104.5	111.8	100.2	101.9	107.2	105.2
S. D.	14.6	13.6	13.2	12.0	14.4	11.9	13.0	13.0	15.6

E. St. Louis

Tables 5.31 to 5.35 present the comparisons for E. St. Louis five-year-old Head Start. The five-year-old Head Start is compared with the 1st-starting Follow Through children. Comparisons at entry to first grade are not available. Seven of the eight differences at the end of third grade favor the Head Start children and three are significant. The levels of significance are as follows:

WRAT		
	Reading	N. S.
	Arithmetic	.001
	Spelling	N. S. (favors 1st-starting Follow Through)
MAT		
	Total Reading	.01
	Total Math	N. S.
	Language	N. S.
	Spelling	.01
SIT IQ		N. S.

The magnitude of the significant differences are on the order of a quarter standard deviation, or from .2 to .4 grade levels.

- - - - - - - - - - -

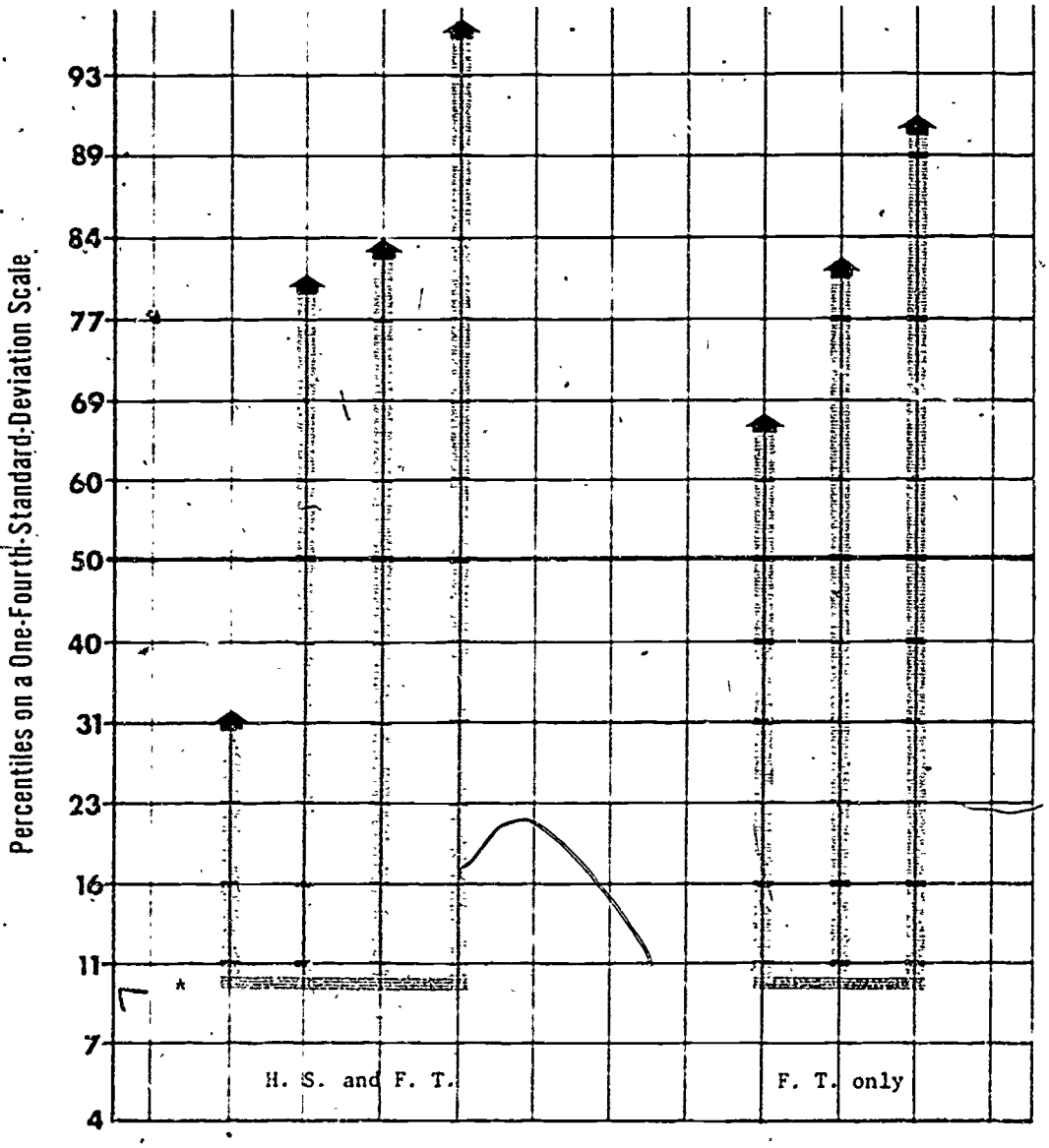
Tables 5.31 to 5.35

- - - - - - - - - - -

Tables 5.36 to 5.40 present the comparisons for the four-year-old Head Start children with the K-starting Follow Through students. While there is a superiority of the Head Start children on entry to kindergarten, there are no significant differences at the end of third grade. We would caution the reader not to conclude that a program for four-year olds is

Table 5.31

E. St. Louis Five-Year Old Head Start and Follow Through Comparison
WRAT Reading

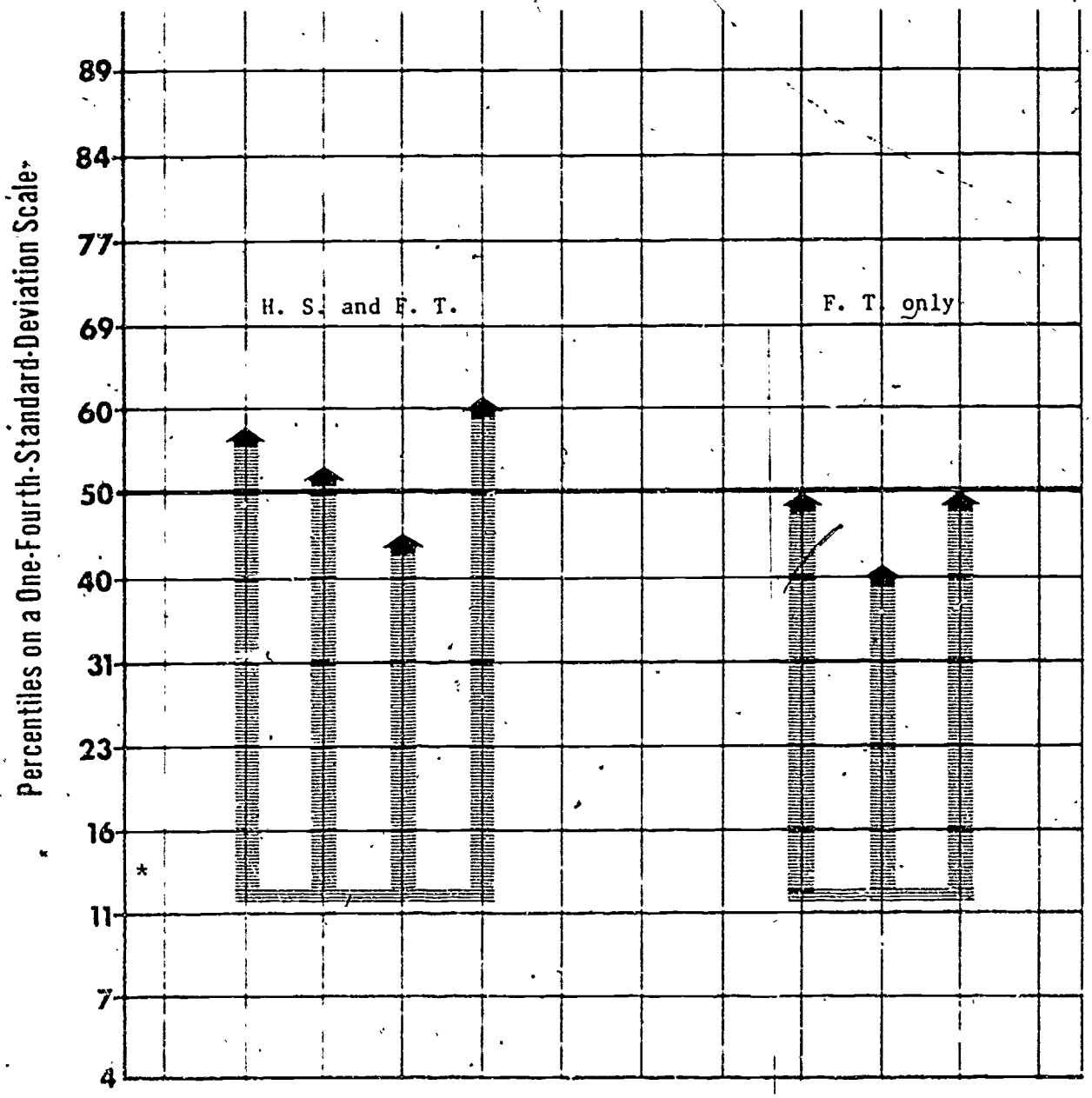


PRE	HS	-	-	-	-	1	-	-	-
POST	-	HS	1	2	3	-	1	2	3
N		50	46	36	37		431	419	389
%-tile		32	81	84	95		68	82	91
S. S.		93	113	115	125		107	114	120
S. D.	No testing	10.8	20.2	16.9	22.6	No testing	21.0	21.3	23.5
G. E.		.72	2.72	4.19	6.46		2.38	4.20	6.05

*E. St. Louis pre-K Follow Through baseline.

Table 5.32

E St. Louis Five-Year Old Head Start and Follow Through Comparison
WRAT Arithmetic

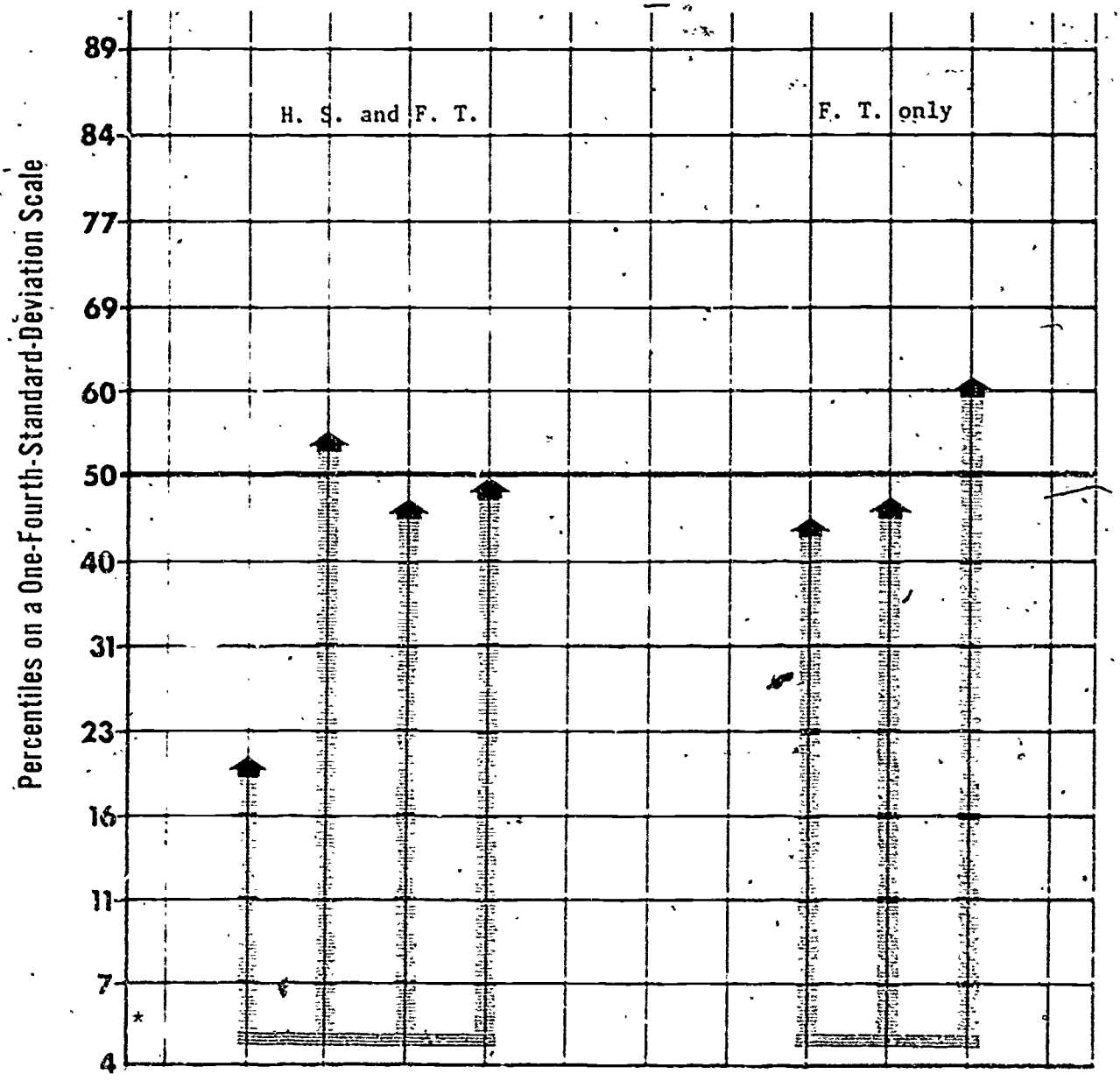


PRE	HS	-	-	-		1	-	-	-
POST	-	HS	1	2	3	-	1	2	3
N	No tests	50	46	36	36	No tests	434	415	388
%-tile		58	53	45	61		59	42	50
S. S.		103	101	98	104		100	97	100
S. D.		12.2	9.5	7.1	7.1		11.7	8.8	10.8
G. E.		1.15	1.99	2.84	4.16		1.91	2.78	3.88

*E. St. Louis pre-K Follow Through baseline.

Table 5.33

E. St. Louis Five-Year Old Head Start and Follow Through Comparison
WRAT Spelling

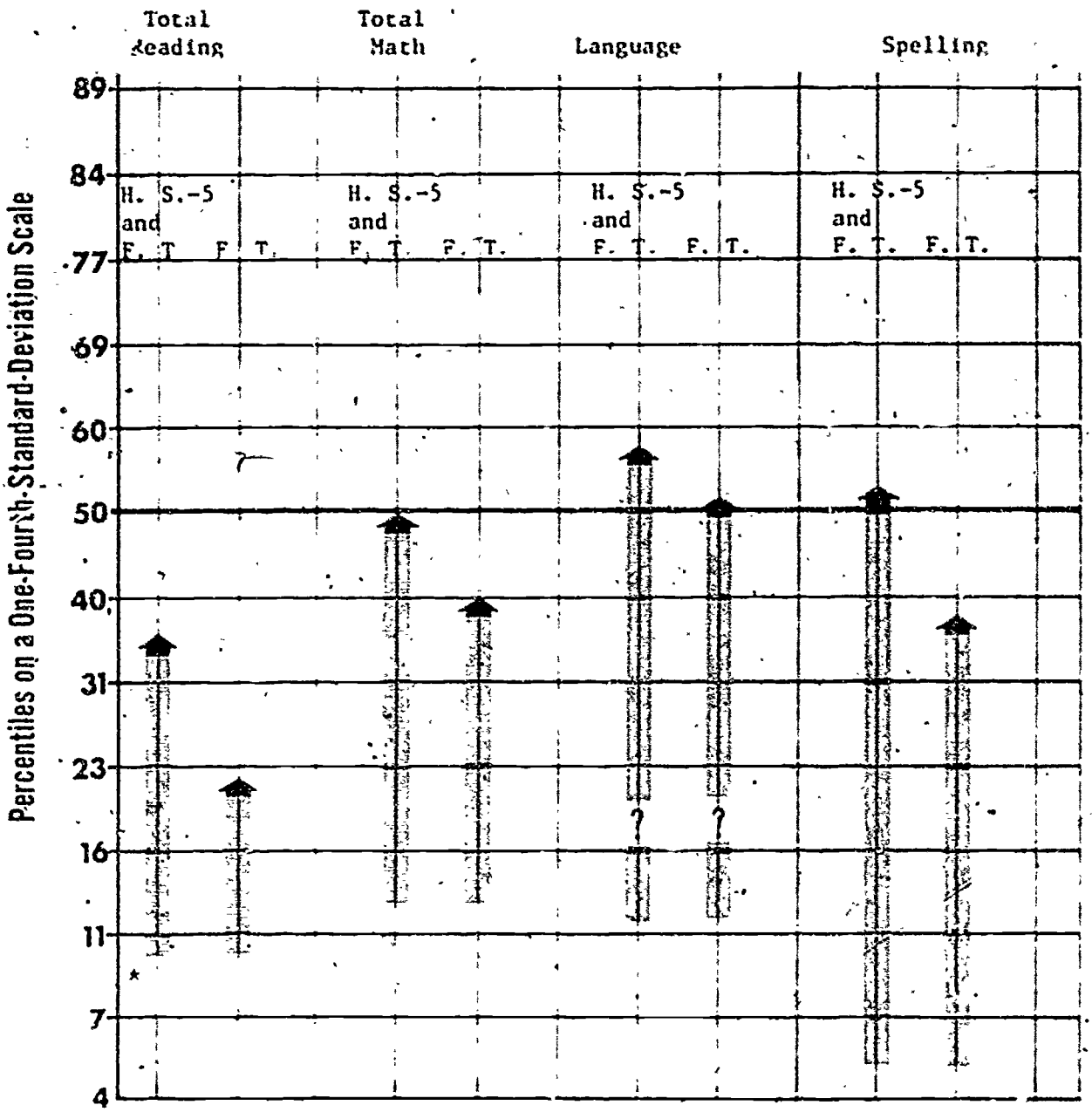


PRE	HS	-	-	-	-	1	-	-	-
POST	-	HS	1	2	3	-	1	2	3
N	No tests	50	46	36	38	No tests	439	420	389
%-tile		21	55	47	50		45	47	61
S. S.		88	102	99	100		98	99	104
S. D.		12.8	10.6	11.7	11.7		12.5	15.1	19.6
G. E.		.51	2.06	2.89	3.75		1.77	2.93	4.25

*E. St. Louis pre-K Follow Through baseline.

Table 5.34

E. St. Louis Five-Year Old Head Start and Follow Through Comparison
MAT

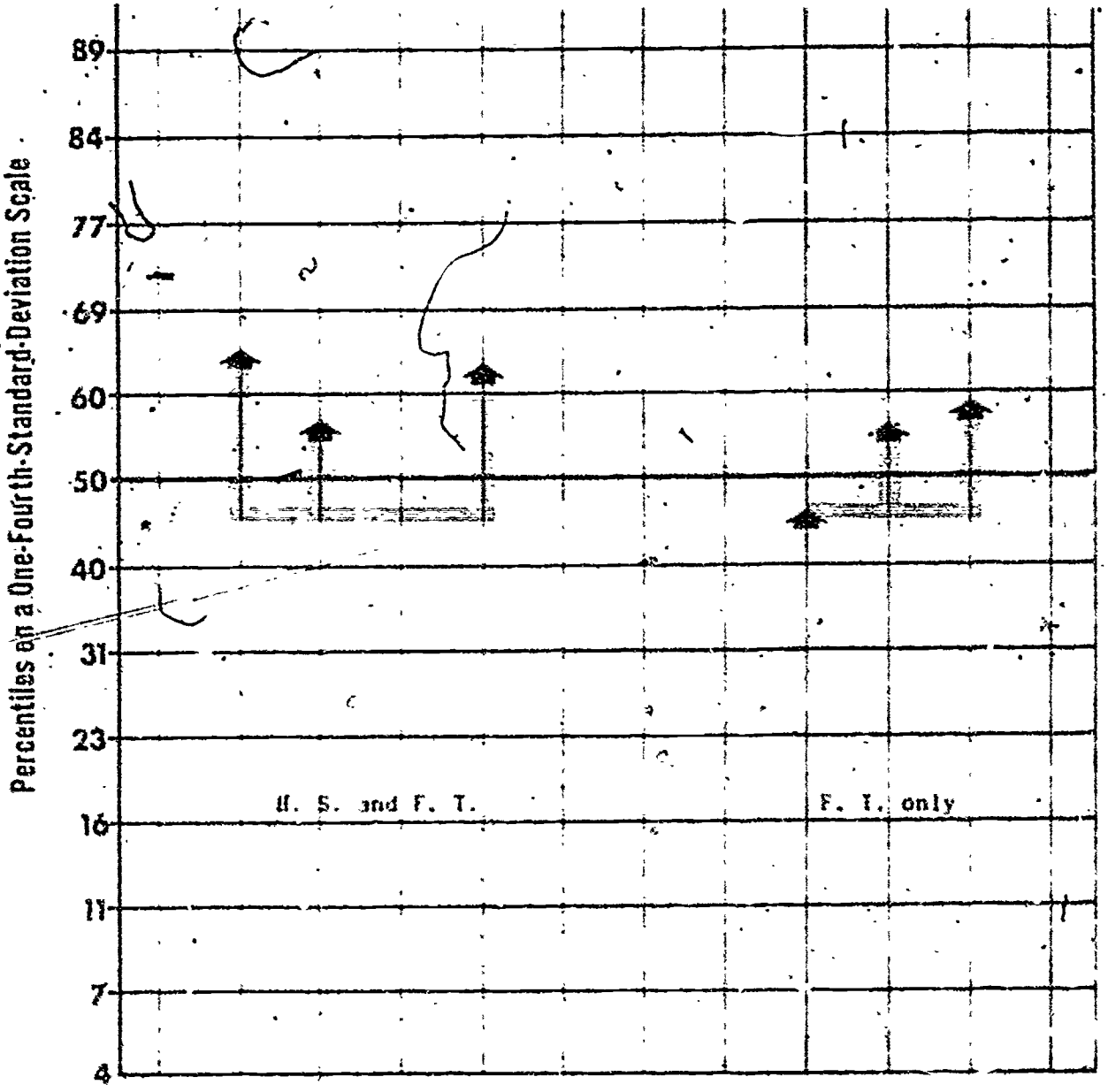


	Total Reading		Total Math		Language		Spelling	
PRE	HS	1	HS	1	HS	1	HS	1
POST	3	3	3	3	3	3	3	3
N	39	106	[illegible]	106	39	103	39	104
%-tile	35	22	50	40	58	52	53	37
S. S.	56.5	52.1	70.4	68.2	72.9	71.6	64.7	58.6
S. D.	7.3	10.5	9.5	11.4	9.4	12.4	10.4	10.8
G. E.	3.11	2.79	3.85	3.69	4.54	4.41	3.87	3.32

*E. St. Louis pre-K, WRAT baseline.

Table 5.35

E. St. Louis Five-Year Old Head Start and Follow Through Comparison
SIT IQ



PRE	85	-	-	-	-	-	-	-	
POST	-	85	1	2	3	-	1	2	3
N		51	43		50		423	121	401
%-tile		65	52		61		47	56	59
S. S.		103.8	102.4		103.1		99.9	102.2	101.3
S. D.		11.5	15.1		14.7		13.9	15.4	15.2
G. E.									

not of value. This program was not well-implemented, aide turnovers and absenses were extremely high, and actual progress in the programs very low. The proper conclusion is that the possible value of a four-year-old program has not been demonstrated.

- - - - - - - - - - -

Tables 5.36 to 5.40

- - - - - - - - - - -

When the five-year-old Head Start program is compared with the public school kindergarten program, the kindergarten children are clearly ahead at the end of the first year as follows:

Standard Scores

	Post HS-5	Post FT-K	p
WRAT			
Reading	93	112	001
Arithmetic	103	112	.001
Spelling	88	109	.001
SIT IQ	105.8	107.9	N. S.

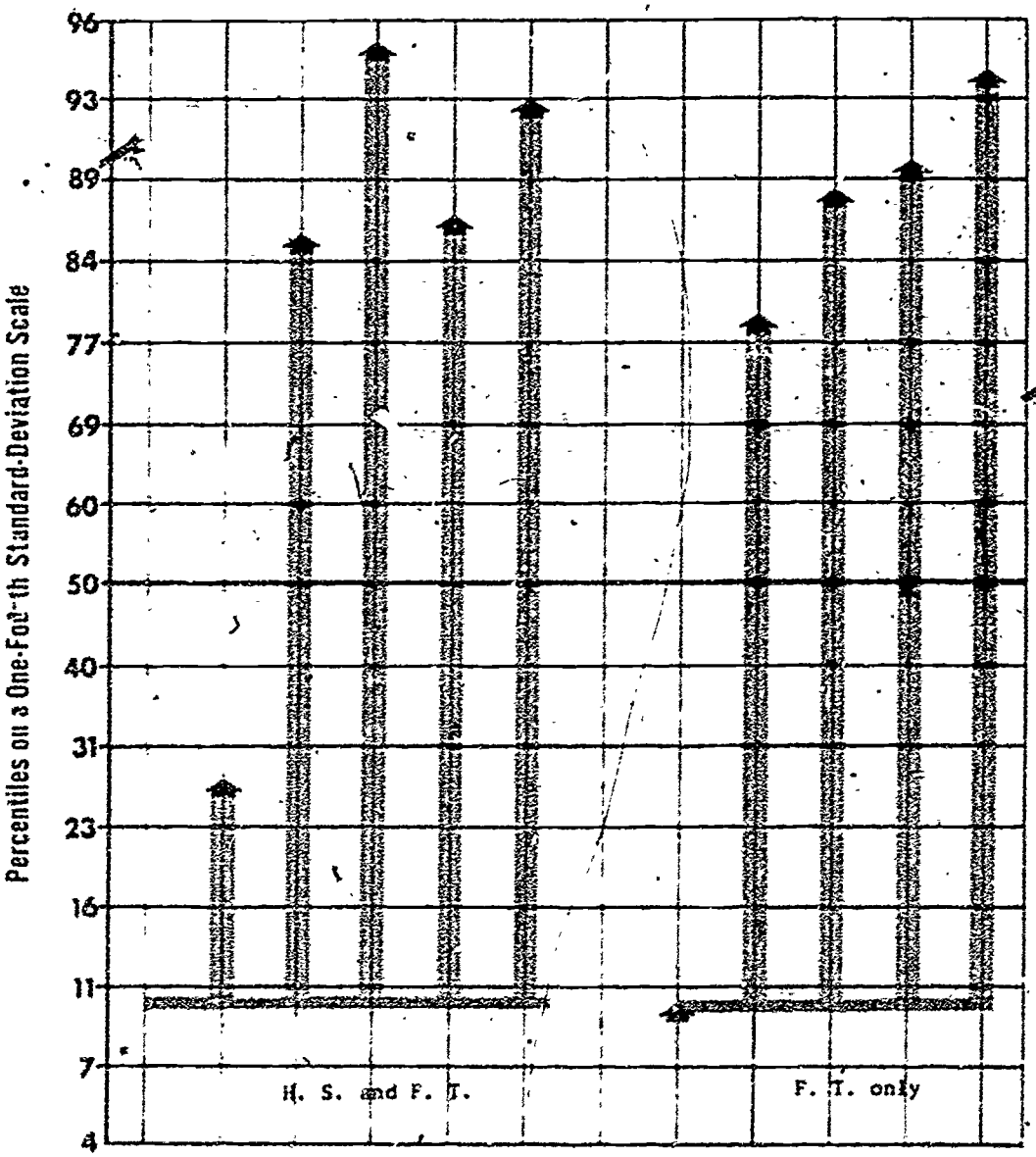
At the end of third grade differences are as follows:

	HS-5	FT-K	Direction	p
WRAT				
Reading	125	123	HS	N. S.
Arithmetic	104	105	FT	N. S.
Spelling	100	106	FT	.05
MAT				
Total Reading	56.5	59.4	FT	N. S.
Total Math	70.4	72.6	FT	N. S.
Language	72.9	71.8	HS	N. S.
Spelling	64.7	63.2	HS	N. S.
SIT IQ	105.1	111.8	FT	.05

Five differences favor the five-year olds taught in Follow Through (two significantly) and three favor the five-year olds taught in Head Start (none significantly). These data give a slight edge to the children taught in the public school setting.

Table 5.36

E. St. Louis Four-Year Old Head Start and Follow Through Comparison
WRAT Reading

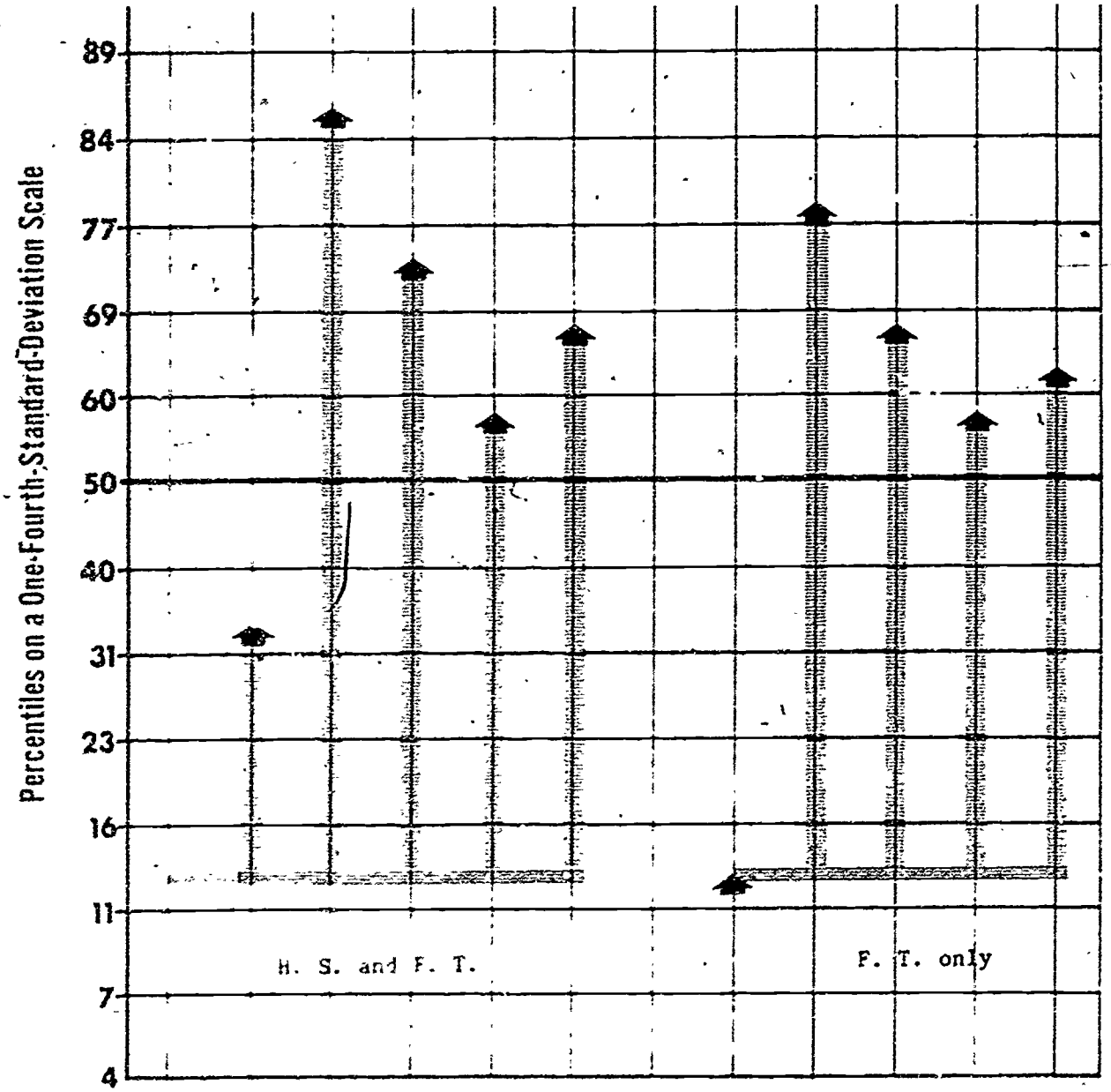


PRE	HS	-	-	-	-	-	K	-	-	-	
POST	-	HS	K	1	2	3	-	K	1	2	3
N	20	44	43	88	77	60	238	610	610	371	160
%-tile	-	27	86	95	87	93	10	79	88	90	94
S S	-	91	116	124	117	122	81	112	118	119	123
S. D.	-	6	21.4	23.4	24.0	24.2	18.8	16.3	22..	22.9	25.4
G. E.	-.57	.44	1.70	3.37	4.[illegible]	6.11	- [illegible]	1.49	3.04	4.72	6.28

*E. St. Louis pre-K Follow Through baseline.

Table 5.37

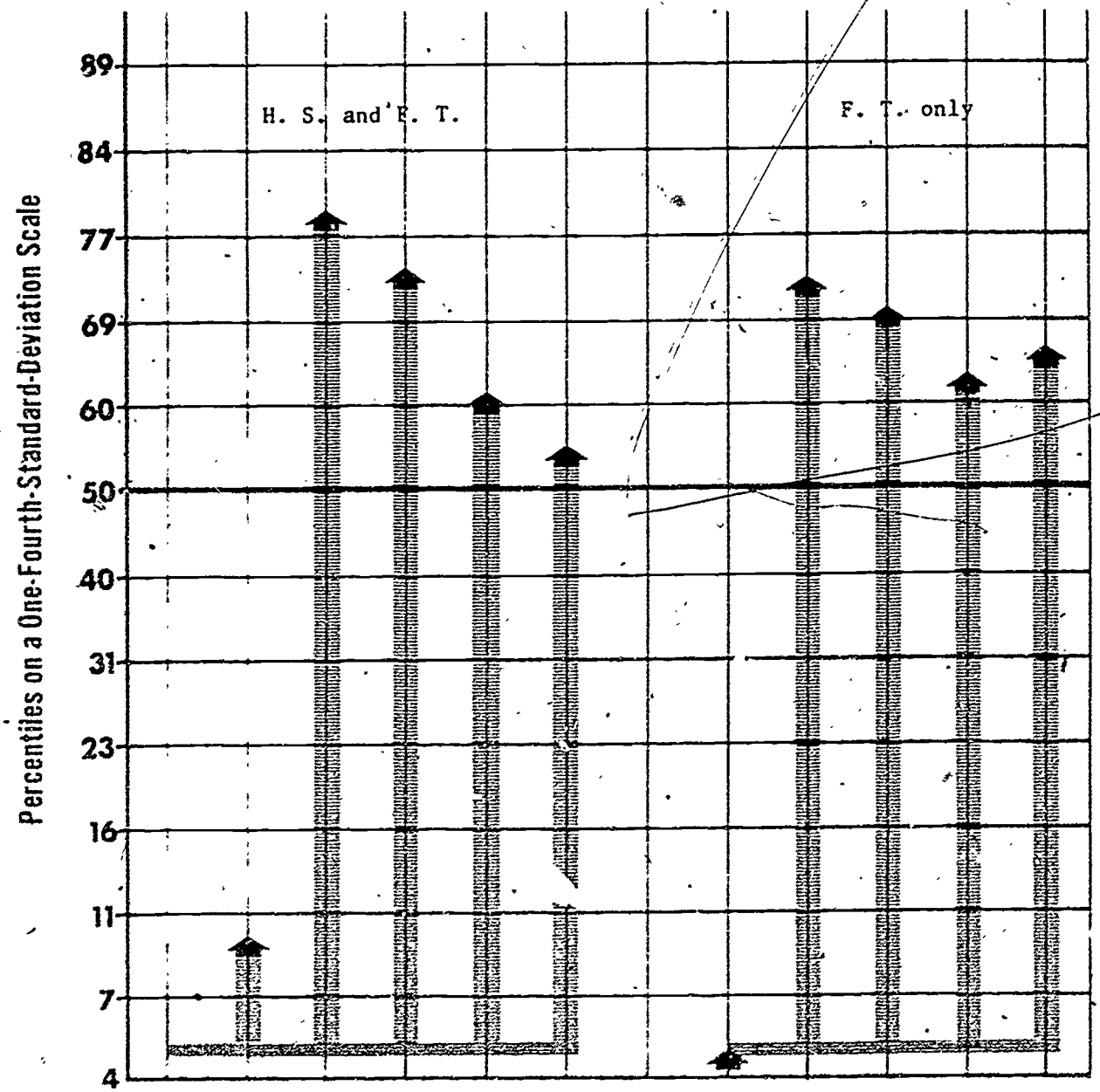
E. St. Louis Four-Year Old Head Start and Follow Through Comparison
WRAT Arithmetic



PRE	HS	-	-	-	-	-	K	-	-	-	-
POST	-	HS	K	1	2	3	-	K	1	2	3
N		44	--	88	76	58	364	608	619	364	159
%-tile			--		58	68	13	79	66	58	63
S. S.		94	116	110	103	107	83	112	107	103	105
S. D.		11.9	16.3	13.1	9.1	6.9	16.0	14.4	12.3	9.6	8.1
G. E.		.55	1.69	2.47	3.20	4.40	.06	1.50	2.34	3.22	4.26

Table 5.38

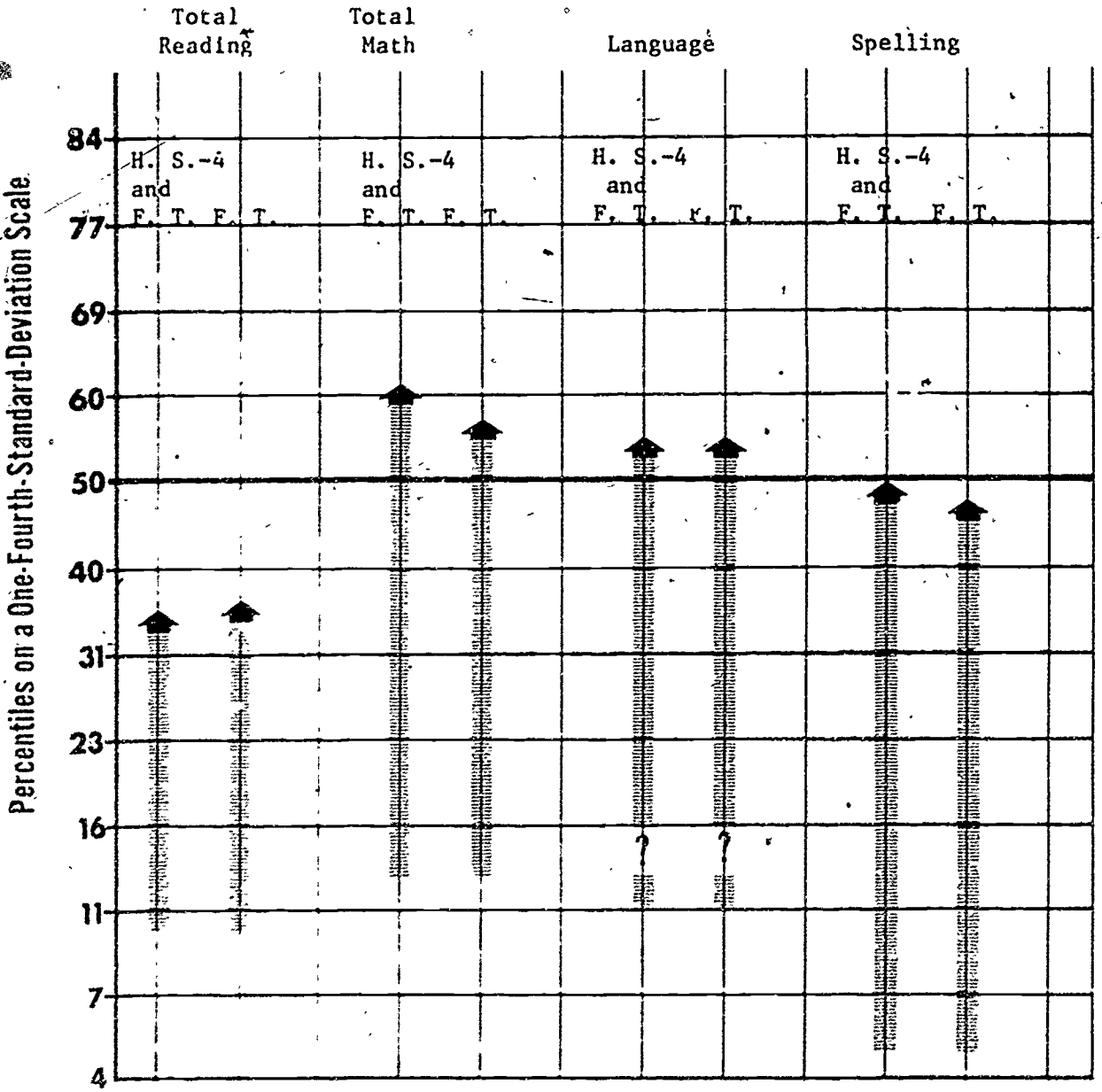
E. St. Louis Four-Year Old Head Start and Follow Through Comparison
WRAT Spelling.



PRE	HS	-	-	-	-	-	K	-	-	-	-
POST	-	HS	Y	1	2	3	-	K	1	2	3
N		43	55	88	78	60	322	632	615	368	160
%-tile		10	79	75	61	55	5	73	70	63	66
S. S.		81	112	110	104	102	76	109	108	105	106
S. D.		18.7	14.7	17.7	16.7	14.5	22.2	16.6	15.0	17.8	19.6
G. E.		-.01	1.53	2.45	3.27	3.86	-.25	1.39	2.36	3.40	4.42

Table 5.39

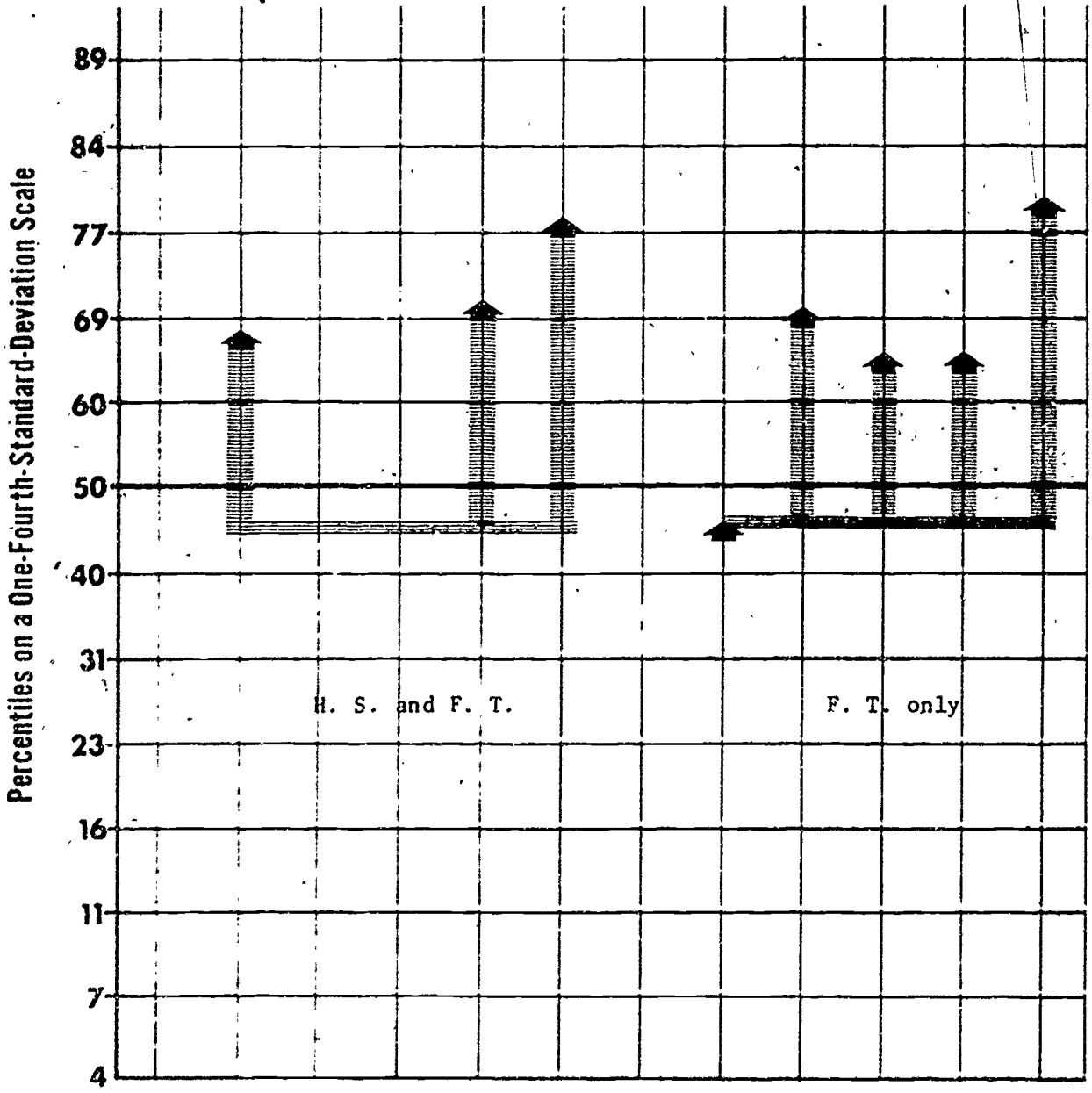
E. St. Louis Four-Year Old Head Start and Follow Through Comparison
MAT



PRE	HS-4	K	HS-4	K	HS-4	K	HS-4	K
POST	3	3	3	3	3	3	3	3
N	60	160	54	159	56	160	56	157
%-tile	36	37	61	57	55	55	50	48
S. S.	37.0	[illegible]	7[illegible]	72.6	71.7	71.8	64.2	63.2
S. D.	7.8	8.8	11.8	9.4	10.4	11.[illegible]	12.2	11.5
G. E.	3.15	3.21	4.20	4.06	4.42	4.50	3.84	3.73

Table 5.40

E. St. Louis Four-Year Old Head Start and Follow Through Comparison
SIT



PRE	HS	-	1	1	1	1	K	-	-	-	-
POST	-	HS	K	1	2	3	-	K	1	2	3
N	-	86	-	-	75	60	483	587	605	268	159
%-tile	-	68	-	-	71	78	46	70	65	65	79
S. S.	-	106.8	-	-	106.1	111.3	98.1	104.9	105.6	105.6	111.8
S. D.	-	13.9	-	-	15.9	16.6	12.7	13.6	14.3	16.5	18.5

Summary on Head Start Comparisons

Significant post-third-grade effects for Planned Variation Head Start children are present for all three sites with five-year olds. An added year of teaching does lead to more proficient kids. Our interpretation of the four-year old Head Start results is that no conclusion is possible because the program was not well-implemented.

A comparison of the magnitude of effects found for the within-site comparisons of 3- and 4-years of program with the effects within site with and without a K-level Head Start is instructive. The average advantage by the end of third grade of the E-B Planned Variation Head Start programs was about one-quarter standard deviation. The average advantage for a 4-year Follow Through Program was about one-half standard deviation. This finding is in keeping with our experience as sponsors that it is much easier to implement and maintain a program within a public school administrative system than within the framework of a Community Action Agency. These findings would support a recommendation to fund Head Start as part of public school programs.

In a later chapter we will present an analysis of the ways in which student attendance, lessons taught, and other process variables relate to outcomes. It is within these analyses that support for the interpretation of the above conclusions about program implementation is found.

CHAPTER 6

FOLLOW UP STUDY OF FIFTH AND SIXTH GRADERS

In the Spring of 1975, fifth and sixth graders were tested with the WRAT (Levels 1 and 2) and the MAT (Intermediate) in seven sites where control groups could be found. In four sites the SIT was also given to be used as a dependent variable. Only the fifth graders in Flippin and Flint had started the program in kindergarten. All other groups started in first grade. Because of this fact, the results to be reported represent a low estimate of what might have been achieved. The only selective factors operating in choosing sites were the availability of a control group and district cooperation.

There were 711 Follow Through students and 766 Non-Follow Through students in the various comparisons. Low-Income students were sought in each case except Flippin, where all students were used because of the small N available and the probable equality of the two communities being compared (Flippin and Cotter). As it turned out, there were a few Non-Low-Income students in most groups, so Low-Income, Non-Low-Income was used as a covariate. Other sample differences in education of parents, sex of student, number of siblings, and Ethnic group (White, Non-White) were also adjusted by the use of covariance analysis where possible.

In Tupelo, Mississippi, the local district California Achievement Test results were used in place of the MAT and the reading scores were

converted to MAT equivalents using information from the Anchor Test Study. In Flint, the test results for MAT Reading were provided to us by the local district for Follow Through and a comparison group.

Covariance analyses were computed for each variable for each site. This provided for a Total of 149 comparisons. Since the group sizes tended to be small (averaging about 50, but ranging from 18 to 117), we accepted a one-tail .05 significance level. We also examined trends with probability levels of .15 (one-tail) or less. Using the .05 level (one-tail test), differences of approximately a quarter standard deviation would be significant when the two sample sizes were 50.

The students tested by site are listed in Table 6.1.

The overall pattern of findings is shown in Table 6.2.

The results show 53 significant differences out of a possible 149 at the .05 level using a one-tailed test. Of these 53, 50 favor Follow Through and 3 favor Non-Follow Through groups. In addition the trends between .06 and .15 probability levels show 19 out of 24 trends in favor of Follow Through. These findings imply that the E-B Follow Through model has produced effects which are still detectable two and three years after the special program is no longer in effect, and under conditions when there was no assurance of a building on the gains made in Follow Through. Furthermore, the effects come primarily from 1st-starting sites where students are typically a quarter to a half standard deviation behind the K-starting sites at the end of third grade.

- - - - - - - - - -

Table 6.1 and 6.2

- - - - - - - - - -

Table 6.1

Sample Sizes and Sites

	Follow Through		Non-Follow Through	
	5th	6th	5th	6th
Flippin, Arkansas	24*	18	57	61
E. St. Louis, Illinois	43	47	45	44
Smithville, Tennessee	47	48	51	47
Uvalde, Texas	117	101	86	101
Dayton, Ohio	104	33	87	47
Tupelo, Mississippi	46	38	34	25
Flint, Michigan	35*	-	82	-
Totals	426	285	441	325

*Started Follow Through in kindergarten.

Table 6.2

Significance Levels (one-tail test) for Fifth and Sixth Grades
FT-NFT Comparisons with Covariate Adjustments

	WRAT				MAT												
	R_1	R_2	A_1	A_2	WK	R	TR	L	Sp	M Comp	M Con	MPS	TM	Sci	FT ++ $p <$ .05	FT + $p =$ .06 .15	0 $p >$ .15	NFT − $p =$ .06 .15	NFT −− $p <$ .05
Flippin, 5th	.06 +	.07 +	.09 −	.06 −	.05 +	.03	.03 +	.10 +	N.S.	.13 −	N.S.	.12 +	N.S.	.001 +	4	4	3	3	0
Flippin, 6th	N.S.	N.S.	N.S.	N.S.	.09 −	N.S.	N.S.	N.S.	.03 −	N.S.	N.S.	N.S.	N.S.	N.S.	0	0	12	1	1
E. St. Louis, 5th	−	−	−	−	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	0	0	10	0	0
E. St. Louis, 6th	−	−	−	−	.01 +	.10 +	.04 +	.08 +	N.S.	.06 +	.10 +	N.S.	N.S.	N.S.	2	4	4	0	0
Smithville, 5th	.01 +	.006 +	N.S.	N.S.	N.S.	.14 +	N.S.	N.S.	.05 +	N.S.	N.S.	N.S.	N.S.	.15 +	3	2	9	0	0
Smithville, 6th	.001 +	.001 +	.001 +	.001 +	.001 +	.001 +	.001 +	.001 +	.006 +	.004 +	.001 +	.001 +	.001 +	.002 +	14	0	0	0	0
Uvalde, 5th	.001 +	.001 +	.02 −	.07 +	N.S.	.07 −	N.S.	N.S.	N.S.	N.S.	N.S.	.01 +	N.S.	.15 +	3	2	7	1	1
Uvalde, 6th	.002 +	.002 +	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	.05 +	.04 +	.02 +	.07 +	.02 +	.14 +	6	2	6	0	0
Dayton, 5th	.001 +	.001 +	N.S.	.001 +	.006 +	.004 +	.003 +	.01 +	.05 +	.03 +	.002 +	.003 +	.003 +	.001 +	13	0	1	0	0
Dayton, 6th	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	.05 −	N.S.	N.S.	.10 +	N.S.	N.S.	0	1	12	0	1
Tupelo, 5th	.01 +	.003 +	N.S.	N.S.	.08 +	N.S.	.13 +	−	−	−	−	−	−	−	2	2	3	0	0
Tupelo, 6th	.005 +	.001 +	.07 +	.05 +	.15 +	N.S.	N.S.	−	−	−	−	−	−	−	3	2	2	0	0
Flint, 5th	−	−	−	−	N.S.	N.S.	N.S.	−	−	−	−	−	−	−	0	0	3	0	0
No. Significant FT+	7/10	7/10	1/10	3/10	4/13	3/13	4/13	2/10	4/10	3/10	3/10	3/10	3/10	3/10	50	19	72	5	3

Examination of the patterns of significance by variable show particularly strong results for Reading. On the WRAT Reading measures, 14 of the 20 comparisons are highly significant in favor of Follow Through. The children have never seen the Level 2 WRAT test before, although they have been tested before on Level 1. On the MAT Reading measures, 11 out of 39 comparisons were significant in favor of Follow Through. Five other reading comparisons for the MAT or WRAT were close to significance ($p < .10$). On WRAT Arithmetic, significant comparisons favoring Follow Through were found in only 4 of 20 instances. The WRAT Level I Arithmetic test appears to be very insensitive to the program as later comparisons will show. On MAT Math measures, 12 of 40 comparisons were significant. The remaining MAT measures showed 3 of 10 significant for Science, 2 of 10 for Language, and 4 of 10 for Spelling. One negative comparison on the WRAT was for WRAT Level I Arithmetic. In this case, the Level 2 Arithmetic test was almost significant in favor of Follow Through ($p < .07$). The two other significant negative findings were on the MAT Spelling test. As we have noted before, Spelling was not a strong point of the program in the early years.

On MAT Math, there is a strong trend for MAT Problem Solving to show significant, or nearly significant, effects even when other measures of math number facts are going against Follow Through groups. This suggests that the problem solving approach taught in Distar Arithmetic is having long-term effects. We would expect even stronger results to show when the students get to algebra, since Distar focuses on row functions to prepare students for what has been traditionally a difficult subject.

The findings in Science are a nice surprise and a strong effect. They can probably be attributed to the Distar 3 Reading Program which uses a science content base to teach students to learn new rules and to apply them in their reading tasks.

MAGNITUDE OF EFFECTS BY SITE

We will examine more carefully the comparability of the FT and NFT groups by site and illustrate the general levels of performance. With small samples, using intact groups, it does not take much in the way of selection bias to produce good or bad results. To allow an examination of group comparability, a Student Background Form was filled out by persons hired at the sites to do this job (and/or by the Follow Through parent workers). For each child the following information was sought:

1. Sex of student.
2. Level of mother's education.
3. Level of father's education.
4. Number of siblings.
5. Income status (above or below OEO guideline). Other evidence accepted to define Low-Income was eligibility for free lunches or eligibility for Aid to Dependent Children.
6. White, Non-White ethnic group status.
7. Age.

Because there were some differences between groups on these measures, up to five of them were used as covariates to adjust mean outcomes for the differences. (The available computer program did not permit the use of more than five covariates.) Education of parents was not available in several sites. Where missing data on a covariate occurred (less than 2 percent of the cases), the group mean for this site was inserted. Age was not used as a covariate, but to determine WRAT standard scores which are age-normed rather than grade-normed.

Flippin

The comparison group for Flippin came from Cotter, a nearby community of similar composition and slightly larger in size. All but one student was White. Table 6.3 shows the means and standard deviations on the background variables for the Flippin comparisons. The Follow Through mothers tend to have more education than the Non-Follow Through mothers. In the sixth grade, the NFT group has more girls, more sibs and the students are three months older. The variables used as covariates are marked with an X.

- - - - -

Table 6.3

- - - - -

The variable influencing covariance adjustments the most for the fifth grade comparisons was Father's Education. The adjustment tended to improve the advantage of Follow Through students. Sex of student, parental education, and income status had minor influences on the adjusted outcomes for the sixth grade comparisons.

Table 6.4 and 6.5 show the unadjusted mean scores and percentile standings. For the WRAT, all calculations were based on raw scores. The norm tables for Level 1 do not go above 12 years old, and those for Level 2 do not go below. This creates a dilemma. Mean raw scores were converted to mean grade equivalents and then to percentiles (using a mean age adjustment). To indicate that these figures are based on a different computation procedure than used previously (or still used for the MAT), they have been placed in parentheses. To reduce the total presentation requirements, only the Level 1 WRAT scores are illustrated along with selected MAT scores. The superiority of the fifth grade

- - - - - - - - - -

Table 6.4 and 6.5

- - - - - - - - -

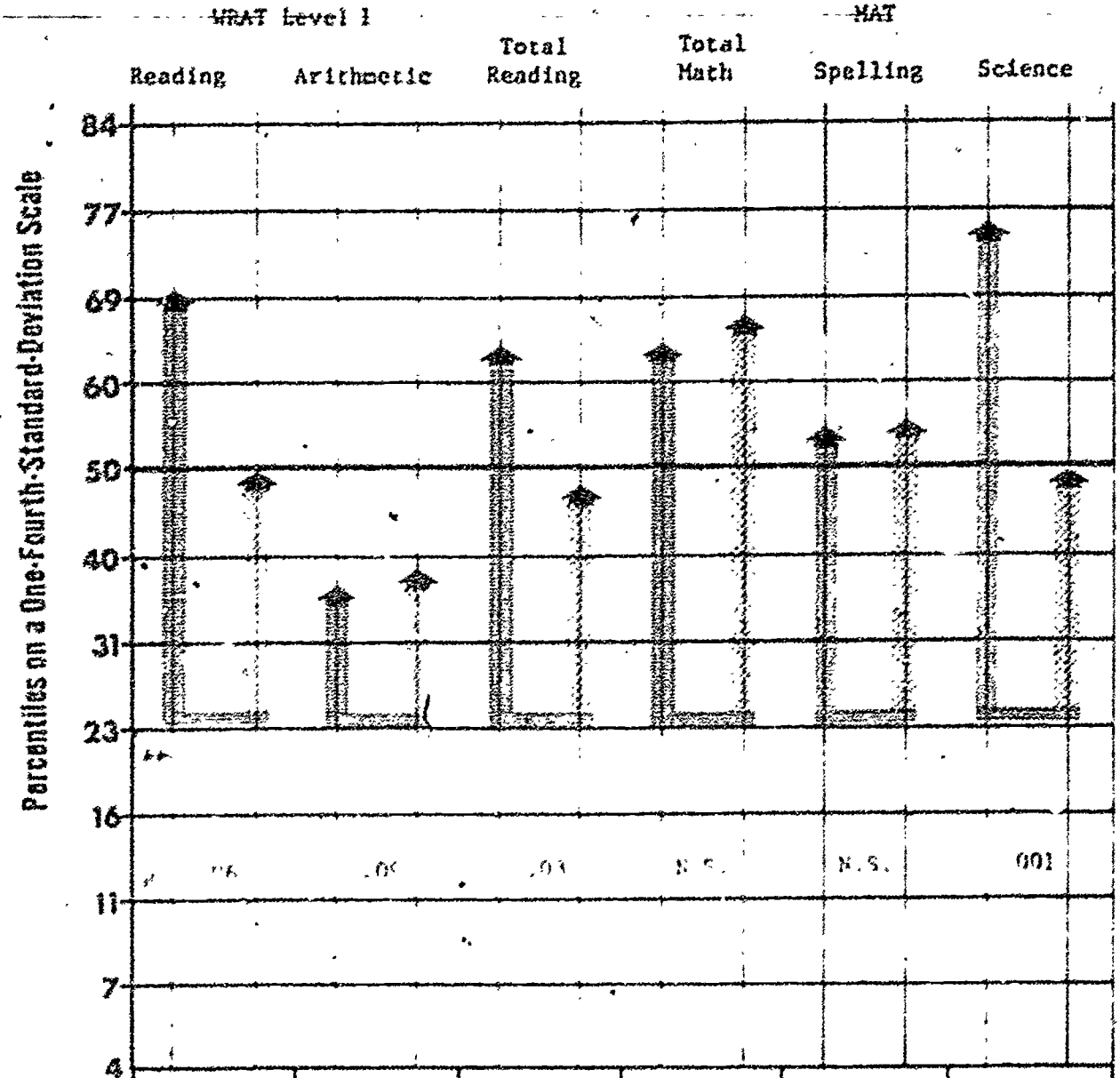
Table 6.3

Background Variables for Flippin FT-NFT Comparisons for Fifth and Sixth Grades

	Fifth				Sixth				
	FT		NFT		FT		NFT		
N	24		57		18		61		
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Used as Covariate (X)
Sex (1 = Male, 2 = Female)	1.5	.51	1.42	.50	1.44	.51	1.57	.50	X
Educ. Mother (4 = High School Grad.)	4.62	1.06	4.45	1.43	4.89	.83	4.30	1.59	X
Educ. Father	4.38	.88	4.56	1.06	4.50	1.34	4.32	1.40	X
No. of Sibs	3.46	1.72	3.14	1.27	3.22	1.80	3.92	2.16	X
Income Level (1 = low) (2 = not low)	1.67	.48	1.64	.48	1.67	.48	1.72	.45	X
Ethnic Group (1 = White) (2 = Non-White)	1.00	.00	1.00	.00	1.00	.00	1.07	.14	
Age of Student (years)	11.1		11.2		12.0		12.3		

Table 6.4

Flippin Fifth Grade FT-NFT Comparisons for Unadjusted Test Scores; p Values are for Covariance Adjusted Scores

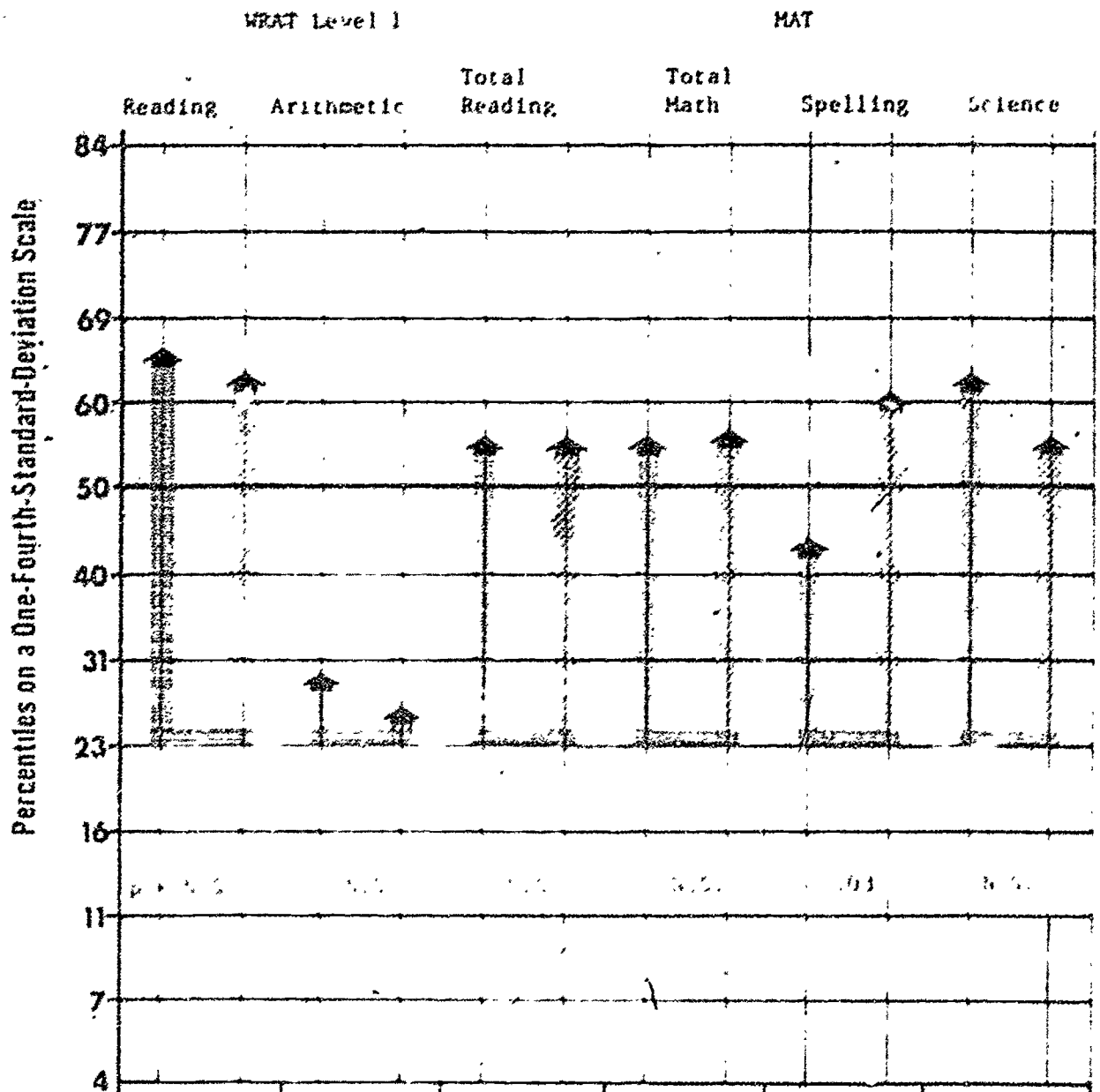


GROUP	FT	NFT	FT	NFT	FT	NFT	FT	NFT	FT	NFT	FT	NFT
POST	5	5	5	5	5	5	5	5	5	5	5	5
N	24	57	24	57	24	57	24	57	24	57	24	57
%-tile	(70)	(50)	(37)	(39)	64	48	64	67	54	55	76	50
S. S.*	24.0	69.6	35.5	35.9	83.0	77.8	90.8	91.3	79.4	79.5	89.0	82.8
S. D.*	9.7	13.3	2.9	5.3	11.3	13.4	7.6	11.2	10.5	12.9	9.5	9.0
G. E.	(7.6)	(6.9)	(5.2)	(5.5)	6.5	5.8	6.2	6.4	5.8	5.7	7.1	6.0

*WRAT scores are raw scores.
**Baseline is arbitrary

Table 6.5

Flippin Sixth Grade FT-NFT Comparisons for Unadjusted Test Scores; p Values are for Covariance Adjusted Scores



GROUP	FT	NFT	FT	NFT	FT	NFT	FT	NFT	FT	NFT	FT	NFT
POST												
N	18	61	18	61	18	61	18	61	18	61	18	61
%-tile	(66)	(63)	(30)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	44	61	[illegible]	[illegible]
S. S.*	22.9	28.1	16.5	18.1	[illegible]	[illegible]	96.6	94.7	82.4	88.1	[illegible]	[illegible]
S. D.*	8.1	9.6	3.8	4.3	[illegible]	[illegible]	[illegible]	10.7	8.4	11.0	[illegible]	[illegible]
G. E.	(7.8)	(7.9)	(5.8)	(5.9)	6.9	6.9	6.8	6.8	6.2	7.0	[illegible]	6.9

*WRAT scores are raw scores.
**Baseline is arbitrary.

Follow Through children is on the order of one-half standard deviation on both Reading measures and in Science. The scores on WRAT Arithmetic are strangely low and quite inconsistent with the above grade level scores for MAT Math. The repetition of this pattern in many tables to come leads us to question the validity of the WRAT Arithmetic test at this level. Table 6.5 shows that except for WRAT Arithmetic and Spelling (our early neglect again), the Follow Through children are above grade norms in the sixth grade, even though the differences do not significantly favor Follow Through.

CIT IQ was also examined as a dependent variable using the same covariates. Mean IQ for the fifth grade Follow Through children was 112.8 and that for the comparison group 106.9. The covariance adjustments for differences in background variables did not change the magnitude of this difference. The probability level for the difference was .09 (one-tailed test). These Follow Through children from Cohort 7 (K-starting) left kindergarten with a mean IQ of 111.2 (N = 71) and left third grade with a mean IQ of 120.5 (N = 71). One might wish to argue that the differences in Reading and Science are due to IQ differences, but that would not explain the failure to find differences in Spelling and Math. We would not use IQ as a covariate unless it was available for both groups on entry to art 01. The drop in IQ from the end of the third grade to fifth grade should be noted.

For the sixth grade children, mean IQ was 117.4 for FT and 109.8 for NFT. The difference was not significant with or without covariance adjustments. These cohort 4 (1st-starting) children left first grade with a mean IQ of 112.1 (N = 19) and third grade with a mean IQ of 120.4 (N = 18). The drop in IQ is similar to that found for the fifth graders.

The changes in norm-group standing from third to fifth grade are shown below:

	Post Third			Post Fifth			Gain (Loss)	
	%-tile	G. E.	N	%-tile	G. E.	N	%-tile	G. E.
WRAT								
Reading	92	5.8	21	(70)	(7.0)	24	-22	1.2
Arithmetic	73	4.7	21	(37)	(5.2)	24	-36	.5
MAT								
Total Reading	70	4.4	21	64	6.5	24	- 6	2.1
Total Math	88	5.4	21	64	6.2	24	-24	.8
Spelling	60	4.1	21	[illegible]	5.8	24	[illegible]	1.7

The only sharp loss in standing for the Flippin children is on WRAT Arithmetic, which as noted above is producing strangely low findings. The loss on MAT Total Math is less dramatic but sufficient enough to suggest that the math skills taught in Follow Through are not being built upon in a systematic way. The lack of a loss on MAT Total Reading can possibly be attributed to the fact that the Flippin children do not for the most part come from homes where there is minimal language stimulation.

There were no MAT Tests for the current sixth graders when they left third grade. On the WRAT Arithmetic a similar percentile loss was found as was noted for fifth graders (70th to 37th percentile). On WRAT Reading the loss was from the 84th percentile to the 66th.

In summary, the significant findings for the Flippin fifth graders appear to be based on a fair comparison, although it is possible that the groups differ on entry IQ. The differences of a half-standard deviation in Reading and Science must be considered educationally significant.

E. St. Louis

The comparison groups for E. St. Louis come from the same schools as the Follow Through students. All students are Blacks. Table 6.6 shows the background variables. Data on education of parents was available for less than half of the parents and so was not used as a covariate. If the means are typical for the total group, it would appear that the Follow Through parents have a little more education. The parents of 91 sixth graders also have one less child on the average. The covariance adjustments for the sixth grade data are in the direction of reducing the Follow Through effect. The correction is based on area where pir[illegible] to follow throughout, so we do not know what might happen if full data on parent education were available.

Table 6.6

[illegible] the results for four MAT measures (SPAT was not given). The fifth graders are far below the norm median in Reading, Math, and Science. They do well in Spelling. The unusually good spelling scores in E. St. Louis may be related to their practice of giving a lot of writing in the lower grades, first copying from the board and then from dictation. But as will be seen, something must also be happening in fourth and fifth grades.

Table 6.6

Background Variables for E. St. Louis FT-NFT Comparisons for Fifth and Sixth Grades

	Fifth				Sixth				
	FT		NFT		FT		NFT		
N	43		45		47		44		Used as Covariate (X)
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
Sex (1 = Male, 2 = Female)	1.42	.50	1.49	.51	1.60	.50	1.50	.51	X
Educ. Mother (4 = High School Grad)*	(4.35	1.21	4.68	1.29)	(4.50	1.30	4.32	1.14)	
Educ. Father*	61	1.26	4.95	1.19)	(4.51	1.40	4.21	1.30)	
No. of Sibs	5.26	2.67	5.56	1.23	4.26	2.05	5.20	2.26	X
Income Level (1 = low, 2 = not low)	1.07	.26	1.02	.15	1.04	.21	1.00	.00	X
Ethnic Group (1 = White, 2 = Non White)	2.00	.00	2.00	.00	2.00	.00	2.00	.00	
Age of Student (years)	11.2		11.2		11.9		12.5		

*Data on less than half the parents

These late-starting cohort 3 children show the following changes from post-third to post-fifth grade:

	Post Third			Post Fifth			Two-Year Gain (Loss)	
	%tile	G. E.	N	%-tile	G. E.	N	%-tile	G. E.
MAT								
Total Reading	24th	3.8	48	17th	4.1	41	-7	+1.1
Total Math	44th	4.7		16th	4.5	44	-28	+.8
Spelling	46th	4.6	46	41st	5.4	39	-5	+1.8

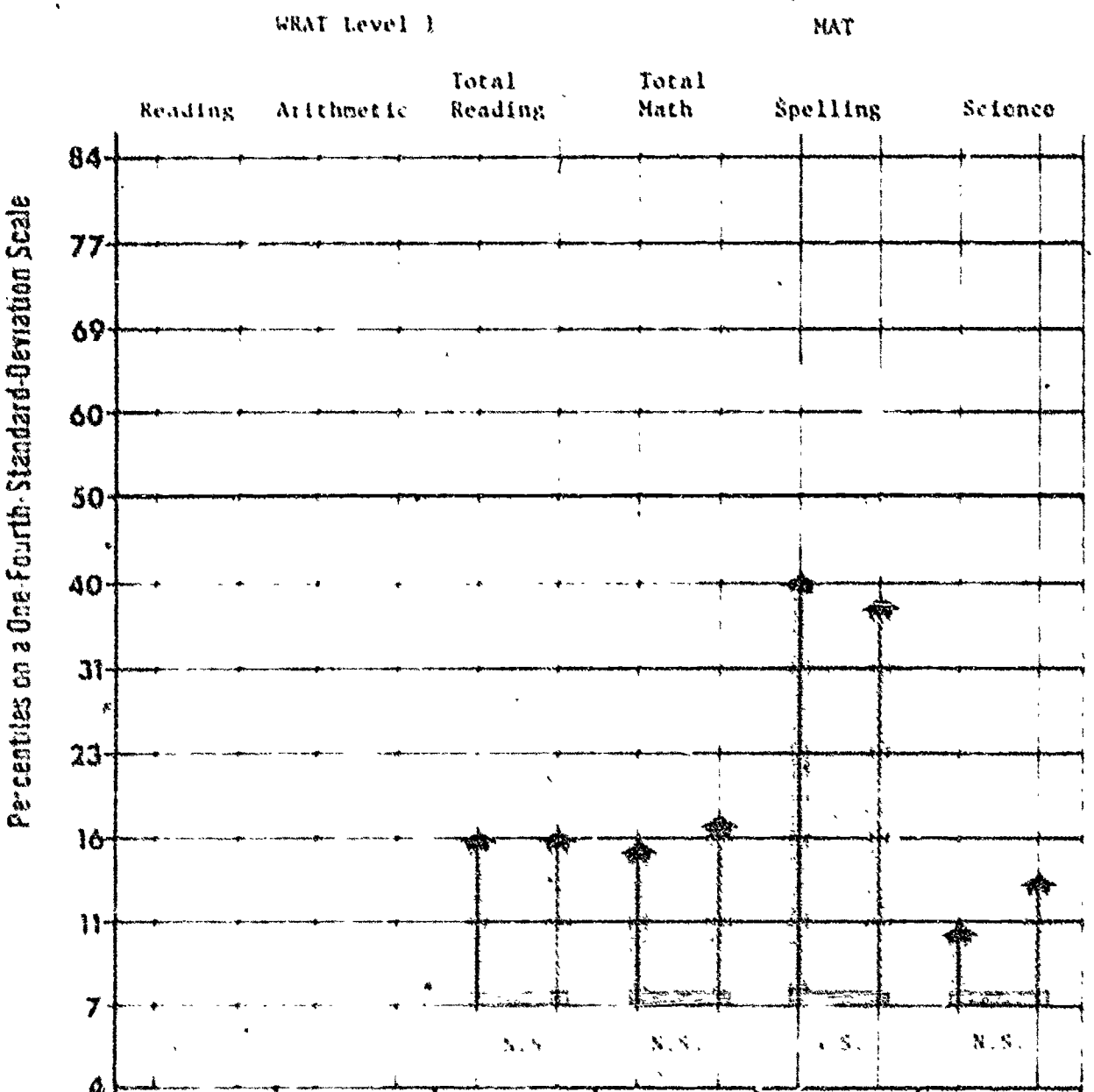
There is a large loss in math. It should be noted that later K starting cohorts leave third grade at much higher levels, as high as the 42nd percentile in Total Reading, 60th percentile in Total Math, and the 59th percentile in Spelling. Thus, the results are not representative of the later progress of children in St. Louis. In any case, the evidence suggests that what happens in the intervening grades can make quite a difference in the outcome of a delayed evaluation of this kind.

Tables 6.7 and 6.8

The comparisons for the sixth grade children from cohort 2 show a higher level of performance for 3 children, exceeding that for NFI by at least a quarter standard deviation in each case. These children completed Follow Through before we started testing with the MAT, so comparisons like that made for cohort 3 are not possible.

Table 6.7

E. St. Louis Fifth Grade FT-NFT Comparisons for Unadjusted Scores; p Values are for Covariance Adjusted Scores

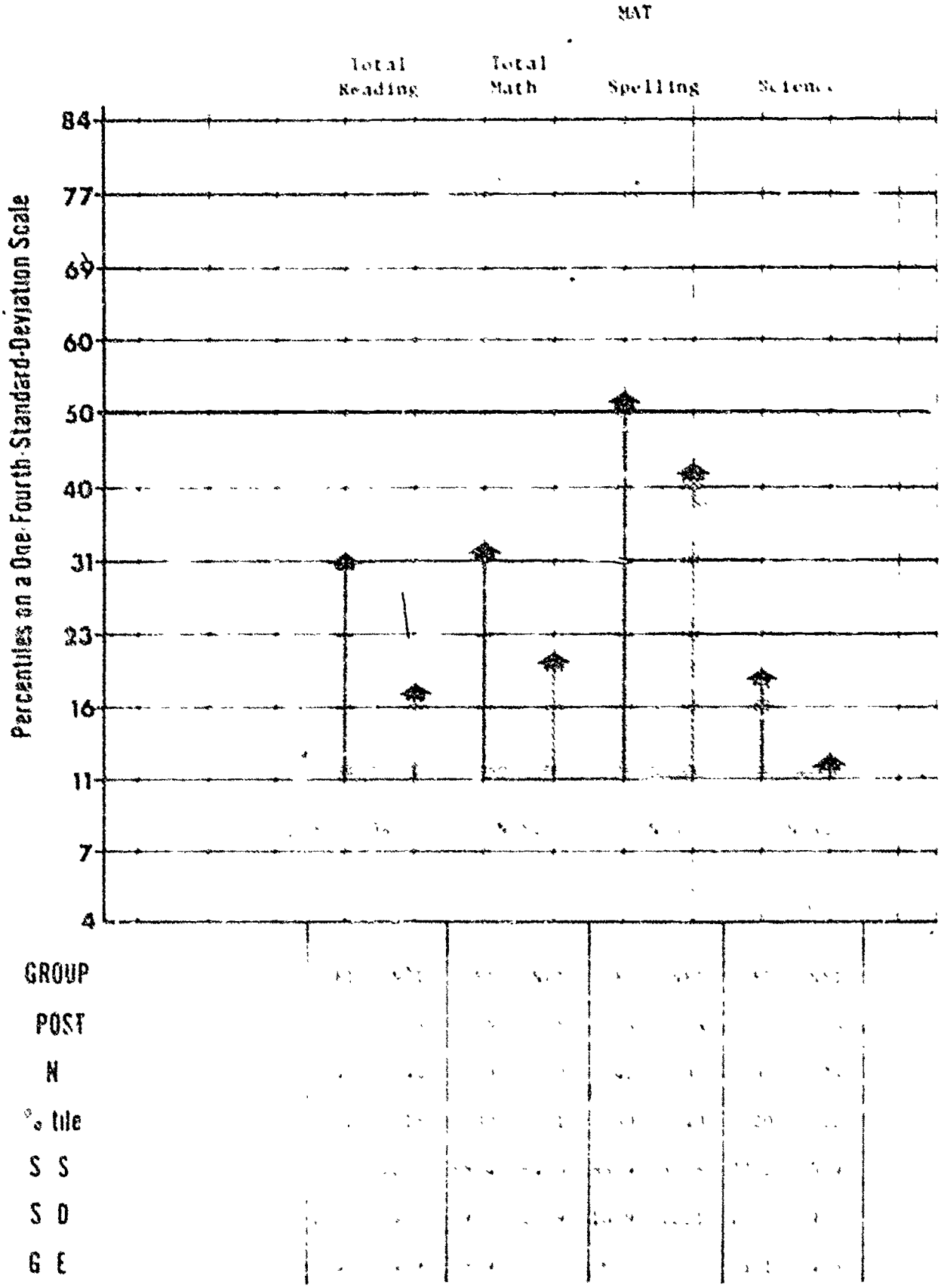


GROUP										
POST										
N										
%-tile										
S. S										
S. D.										
G E										

*Baseline is arbitrary.

Table 6.8

E. St. Louis Sixth Grade FT-NFT Comparisons for Unadjusted Scores; p Values are for Covariance Adjusted Scores



*Baseline to adjusted

The results give some support for positive Follow Through effects for the sixth grade group, and suggest that Follow Through students are losing against the norm group in fourth and fifth grade Reading and Math. Earlier gains against the norm group are being lost. (We will be interested in seeing whether this trend holds for K-starting cohorts in E. St. Louis who reach fifth grade this year.)

Smithville

The comparison students in Smithville come from the same schools as the Follow Through students. The students in both groups are largely Whites. Table 6.9 shows the background measures and indicates the ones used as covariates. In the fifth grade there are more NFT girls, and in the sixth grade more FT girls. The NFT groups also have fewer Low-Income parents than the FT groups. These kinds of inequalities can have drastic effects on comparisons. In the covariance adjustments, sex, number of sibs, and income status have the most effect on the outcomes. Most of the adjustments are directly favoring Follow Through.

Table 6.10 shows the magnitudes of effects for fifth graders. Only WRAT Reading clearly favors Follow Through in the unadjusted data. With adjustments, MAT Spelling also becomes significant. The FT level of performance on MAT Math is at the nation median, even though it is behind the NFT mean in the unadjusted data. With covariance adjustment the situation reverses however. These cohort 3 students show the following changes from post-third to post-fifth grades:

- - - - - - - - - -

Tables 6.9 and 6.10

- - - - - - - - - -

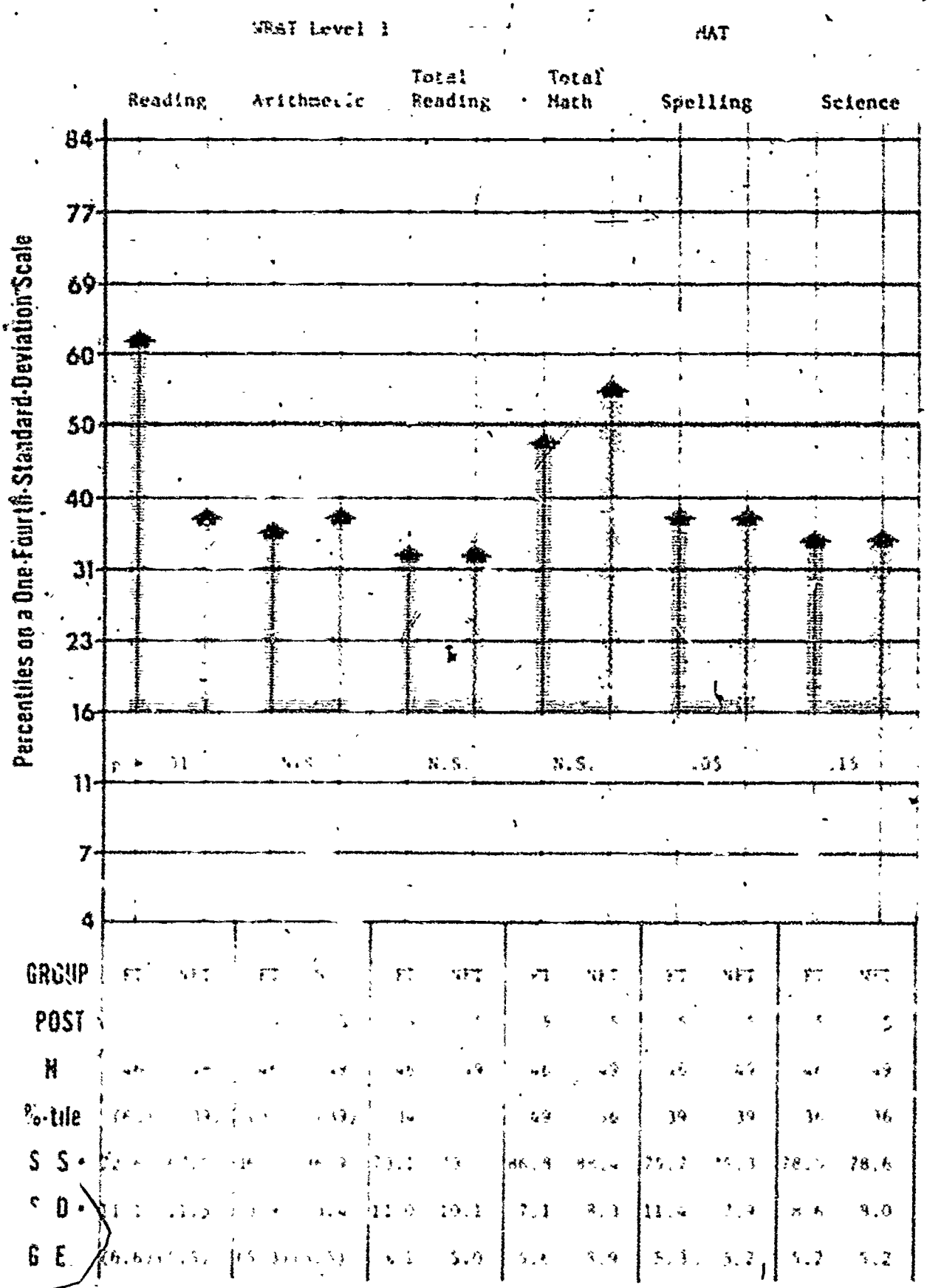
Table 6.9

Background Variables for Smithville FT-NFT Comparisons for Fifth and Sixth Grades

	Fifth				Sixth				
	FT		NFT		FT		NFT		
N	47		51		48		47		
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Used as Covariate (X)
Sex (1 = Male, 2 = Female)	1.49	.50	1.65	.48	1.75	.44	1.51	.50	X
Educ. Mother (4 = High School Grad.)	3.30	1.33	3.70	1.63	3.30	1.30	3.74	1.47	X
Educ. Father	3.72	1.21	4.15	1.12	3.79	1.32	3.87	1.12	X
No. of Sibs	4.15	2.54	4.04	2.43	3.96	2.26	3.91	2.41	X
Income Level (1 = low, 2 = not low)	1.13	.34	1.46	.50	1.12	.33	1.45	.50	X
Ethnic Group (1 = White, 2 = Non-White)	1.17	.37	1.00	00	1.17	.37	1.08	.27	
Age of Student (years)	11.2		11.2		12.0		12.2		

Table 6.10

Smithville Fifth Grade FT-NFT Comparisons for Unadjusted Test Scores; p Values are for Covariance-Adjusted Scores



GROUP	FT	NFT	FT	NFT	FT	NFT	FT	NFT	FT	NFT	FT	NFT
POST							5	5	4	4	5	5
N	49	47	49	49	49	49	49	49	49	49	49	49
%-tile	[illegible]	33	[illegible]	[illegible]	34		69	56	39	39	34	36
S S*	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	86.8	[illegible]	77.2	[illegible]	78.[illegible]	78.6
S D*	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	10.1	7.1	8.3	11.4	[illegible]	[illegible]	9.0
G E	[illegible]	[illegible]	[illegible]	[illegible]	5.1	5.0	5.4	5.9	5.3	5.2	5.2	5.2

*WRAT scores are raw scores.

	Post Third			Post Fifth			Two Year Gain (Loss)	
	%-tile	G. E.	N	%-tile	G. E.	N	%-tile	G. E.
WRAT								
Reading	75th	4.9	46	(63rd)	(6.6)	46	-12	2.3
Arithmetic	67th	4.5	46	(37th)	(5.3)	46	-30	.8
MAT								
Total Reading	52nd	3.6	46	34th	5.1	46	-18	1.5
Total Math	78th	4.9	46	49th	5.6	46	-29	.7
Spelling	62nd	4.3	46	39th	5.3	46	-23	1.0

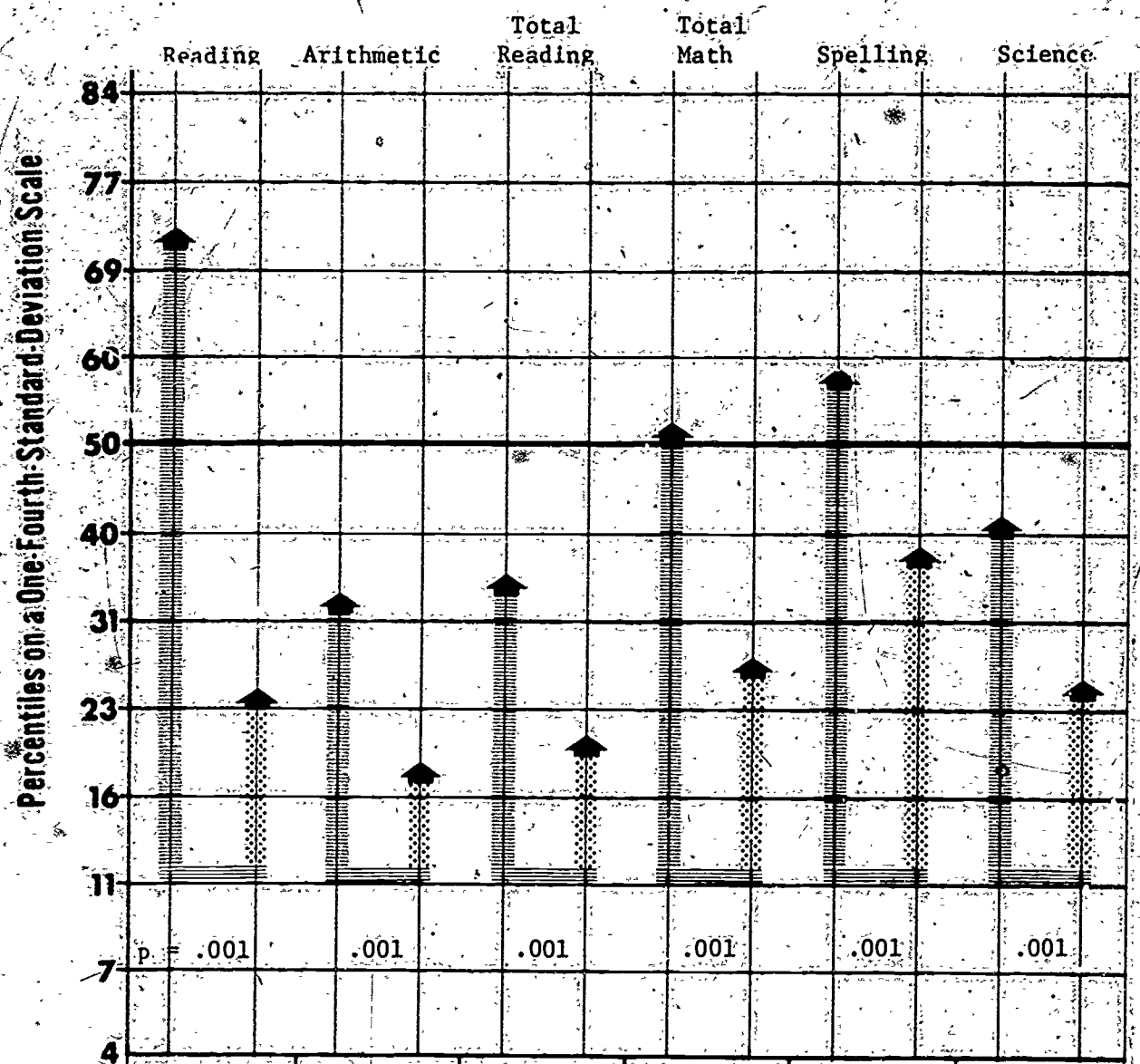
The high performance in Math is most likely a carryover from the lower grades.

The mean SIT IQ for the Follow Through children at the end of fifth grade was 105.1 and that for the comparison group 102.9. The difference is not significant. These cohort 3 students started first grade with a mean IQ of 100.1 ($N = 40$) and left third grade with a mean IQ of 113.4 ($N = 46$). We would have to conclude that these are legitimate comparisons which show a superiority for Follow Through children on WRAT Reading, but losses against national norms during fourth and fifth grades.

Table 6.11 gives the sixth grade comparisons all of which significantly favor Follow Through children. The covariance adjustments operated to increase the differences because the NFT group had more education of parents, higher income, and fewer siblings. The Follow Through students also are significantly higher on SIT IQ before and after covariance adjustment. The mean IQ for FT kids at the end of sixth grade was 108.6 ($N = 48$) and that for NFT, 96.0 ($N = 47$). The Follow Through children had a mean IQ of 102.6 ($N = 49$) at the end of first grade and 115 ($N = 48$) at the end of

Table 6.11

Smithville Sixth Grade FT-NFT Comparisons for Unadjusted Test Scores; p Values are for Covariance Adjusted Scores



GROUP	FT	NFT	FT	NFT	FT	NFT	FT	NFT	FT	NFT	FT	NFT
POST	6	6	6	6	6	6	6	6	6	6	6	6
N	48	45	48	46	48	46	48	45	48	46	48	46
%-tile	(73)	(25)	(34)	(19)	36	21	52	28	59	39	42	26
S. S.*	79.7	68.7	39.9	36.7	80.1	73.5	94.2	87.1	87.3	81.3	84.8	80.1
S. D.*	8.0	11.4	5.9	4.3	11.3	12.1	11.6	9.5	11.2	12.6	9.6	10.4
G. E.	(8.3)	(5.8)	(6.1)	(5.4)	6.1	5.2	6.7	5.8	7.0	6.1	6.3	5.5

*WRAT scores are raw scores.

third grade. Thus, there is reason to believe the difference in IQ at sixth grade may be in part a function of the program. Cohort 2 was not given the MAT in third grade. However, on the WRAT these students left third grade with a mean percentile of 86 in Reading and 68 in Arithmetic. They appear to be maintaining their reading decoding skills. The loss in Arithmetic is probably in part a function of the WRAT test, which is typically producing low fifth and sixth grade performances relative to MAT Math. The Smithville FT sixth graders are still at the 50th percentile on MAT Math.

We conclude that there is a good possibility that the comparison group, even though more advantaged in some ways, has more low performing children in it which may be contributing to some of the differences in favor of Follow Through. However, the performance of the sixth graders from FT is excellent and consistent with their performance on leaving third grade. The overall impact is that of a positive conclusion favoring Follow Through.

Uvalde

Uvalde is a Texas community near the Mexican border. The Follow Through students are mostly Mexican-Americans. It was not possible to find comparison groups in Uvalde who were not in Follow Through. However, comparison groups were found in three communities within 20 miles of Uvalde which have students of similar backgrounds (La Pryor, Batesville, and Sabinal). Table 6.12 gives the information on background variables. The average education of all parents is below the 10th grade level. The groups appear to be well matched on all background variables except that

the fifth grade Follow Through students parents have slightly higher educational levels. Compared to other FT and NFT students, the Uvalde FT and NFT are four months older in fifth grade and eight months older in sixth grade. The covariance adjustments produce only very minor changes in mean effects.

Table 6.12

The results for fifth grade show the 1st-starting cohort 3 FT children ahead on WRAT reading. They are behind significantly on WRAT Arithmetic (Level 1 but not Level 2) which has shown no sensitivity to program effects so far. At the same time they are ahead on MAT Math (significantly, for Math Problem Solving). The general level of performance is on the low side. The mean SIT IQ for FT children (85.9) at the end of fifth grade is significantly higher ($p < .001$) than for NFT (79.6), with or without covariance adjustments. These FT children entered first grade with a mean IQ of 87.4 ($N = 85$) and left third grade with a mean of 95.8 ($N = 117$). In our judgment, the net difference in IQ still favoring the FT group could be a program effect. However, the evidence is open to question.

Table 6.13

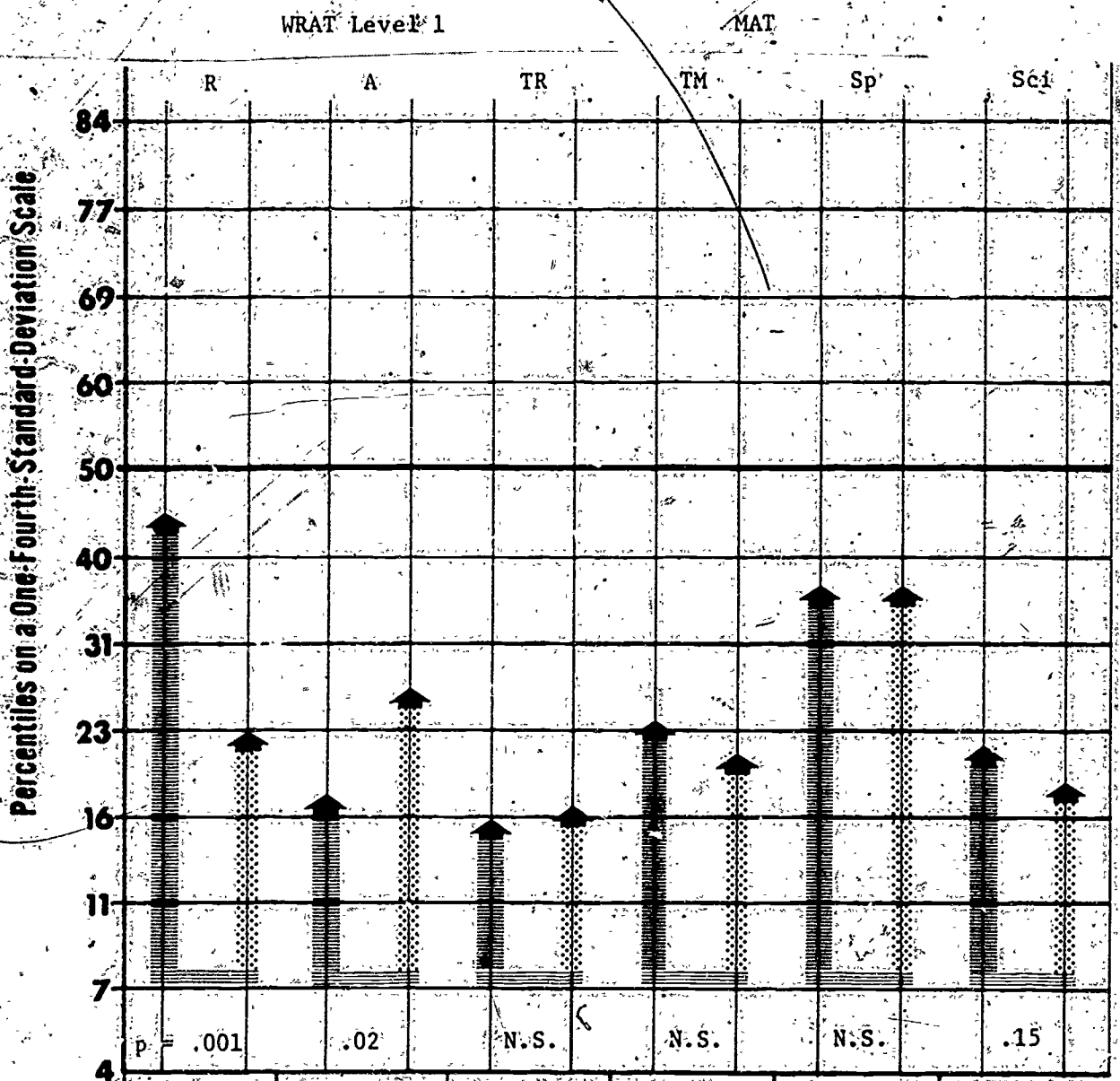
Table 6.12

Background Variables for Uvalde FT-NFT Comparisons for Fifth and Sixth Grades

	Fifth				Sixth				
	FT		NFT		FT		NFT		
N	117		86		101		101		
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Used as Covariate (X)
Sex (1 = Male, 2 = Female)	1.53	.50	1.50	.50	1.52	.50	1.54	.50	
Educ. Mother (4 = High School Grad.)	2.86	1.61	2.34	1.25	2.48	1.10	2.50	1.35	X
Educ. Father	2.86	1.69	2.44	1.21	2.42	1.29	2.36	1.11	
No. of Sibs	6.21	2.87	6.37	2.58	6.34	2.64	6.00	2.48	X
Income Level (1 = low, 2 = not low)	1.13	.34	1.05	.22	1.07	.26	1.02	.14	X
Ethnic Group (1 = White, 2 = Non-White)	2.00	00	1.95	.22	1.99	.10	2.00	00	X
Age of Student (years)	11.5		11.7		12.8		12.8		

Table 6.13

Uvalde Fifth Grade FT-NFT Comparison for Unadjusted Test Scores; p Values are for Covariance Adjusted Scores



GROUP	FT	NFT	FT	NFT	FT	NFT	FT	NFT	FT	NFT	FT	NFT
POST	5	5	5	5	5	5	5	5	5	5	5	5
N	117	83	117	85	117	86	115	82	116	82	116	85
%-tile	(45)	(23)	(18)	(27)	16	17	24	21	37	37	22	19
S.S.*	69.3	63.8	32.1	33.0	64.8	65.4	79.4	78.6	74.8	74.7	74.0	72.4
S.D.*	10.2	12.0	3.2	3.4	9.7	9.5	8.2	7.1	8.5	9.3	7.8	8.0
G.E.	(6.0)	(5.0)	(4.4)	(4.8)	4.0	4.1	4.7	4.6	5.1	5.1	4.5	4.3

*WRAT scores are raw scores.

The changes in performance from third to fifth grade are as follows:

	Third			Fifth			Two-Year Gain (Loss)	
	%-tile	G. E.	N	%-tile	G. E.	N	%-tile	G. E.
WRAT								
Reading	56th	4.6	116	45th	6.0	117	-11	1.5
Arithmetic	39th	3.8	117	18th	4.4	117	-21	.6
MAT								
Total Reading	31st	3.0	117	16th	4.0	117	-15	1.0
Total Math	53rd	3.9	116	24th	4.7	115	-9	.8
Spelling	40th.	3.4	116	37th	5.1	116	- 3	1.7

These changes imply that performance levels in Math are not being maintained after Follow Through, while Spelling is being maintained and taught. The low scores on MAT Reading are consistent with the restriction on English vocabulary levels for children from Spanish speaking homes. As we have pointed out elsewhere (and will again in the Discussion chapter), the vocabulary skills required for performance on MAT Reading, from the Elementary Test Level on, are skills which are for the most part not taught in school, but in the home. The higher relative performance on WRAT Reading is consistent with the initial focus of Distar on decoding skills. The students have learned to decode words as well as middle-class students; they just don't have the breadth of vocabulary comprehension assumed by the tests.

The sixth grade comparisons are shown in Table 6.14. Significant differences favoring Follow Through are present for WRAT Reading measures and most MAT measures except Total Reading. The low WRAT Arithmetic

scores look like those we have seen before. The low Total Reading on the MAT is consistent with our interpretation of the role home language training (vocabulary) plays in learning the skills measured by the test.

A significant difference ($p < .001$) on the Slosson is again found, favoring Follow Through students (84.9 versus 78.3). These FT students entered FT with a mean IQ of 91.4 ($N = 57$) and left third grade with a mean of 93.6 ($N = 92$). Again, we are inclined to interpret the difference at sixth grade as a program effect, although this is not proven.

The Cohort 2 FT students left third grade with a WRAT Reading percentile of 66 ($N = 92$), and Arithmetic percentile of 38 ($N = 93$). By the end of sixth grade, the mean percentiles have dropped to 32 and 14, respectively.

In summary, the Uvalde data show Follow Through strengths in reading decoding and the general competence measured by the SIT. For the sixth graders only, they also show strengths on a variety of MAT skills with the exception of Total Reading.

Table 6.14

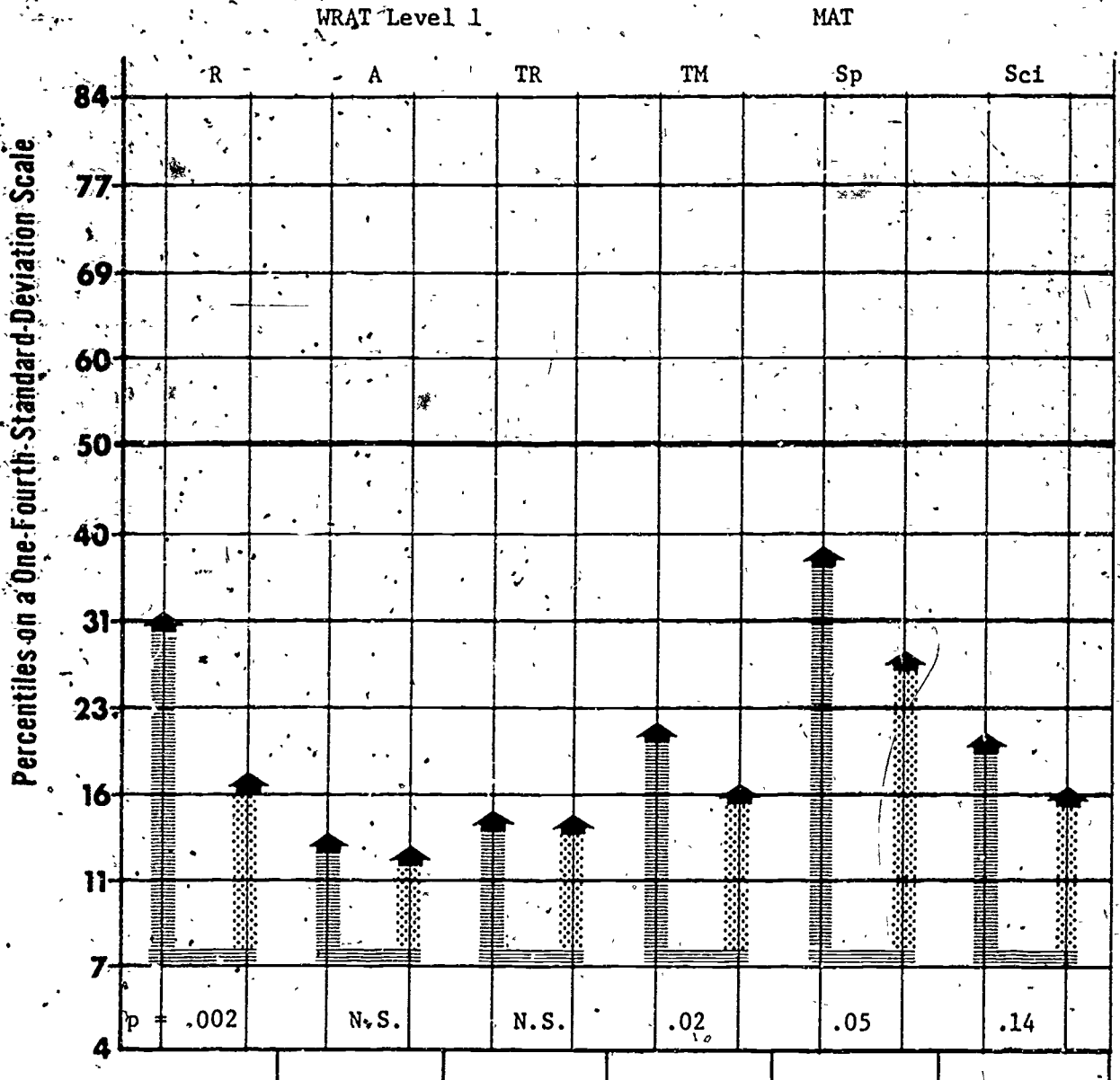
Dayton

The comparison students for Dayton came from a variety of schools within the city. Nearly all FT and NFT students are Blacks. The fifth graders appear to be quite comparable on all variables including the partial information on parental education. The sixth grade background variables appear to favor the NFT group (many more girls, higher education of parents). However, since education of parent could not be used as a covariate due to missing information, the actual adjustments were inconsequential.

Table 6.15

Table 6.14

Uvalde Sixth Grade FT-NFT Comparisons for Unadjusted Test Scores; p Values are for Covariance Adjusted Scores



GROUP	FT	NFT	FT	NFT	FT	NFT	FT	NFT	FT	NFT	FT	NFT
POST	6	6	6	6	6	6	6	6	6	6	6	6
N	101	99	101	100	99	100	98	99	98	99	98	99
%-tile	(32)	(18)	(14)	(13)	15	15	22	17	39	28	21	17
S. S.	*72.7	67.7	34.9	34.5	69.8	70.4	85.0	82.6	81.5	78.5	77.4	76.4
S. D.	*10.6	13.2	4.0	3.9	11.3	10.5	8.1	9.1	11.2	13.2	7.9	8.5
G. E.	(6.6)	(5.6)	(5.2)	(5.1)	4.6	4.7	5.4	5.1	6.1	5.7	5.1	4.9

*WRAT scores are raw scores.

Table 6.15

Background Variables for Dayton FT-NFT Comparisons for Fifth and Sixth Grades

	Fifth				Sixth				
	FT		NFT		FT		NFT		
N	104		87		33		47		Used ad Covariate (X)
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
Sex (1 = Male, 2 = Female)	1.46	.50	1.49	.50	1.36	.49	1.62	.49	X
Educ. Mother* (4 = High School Grad.)	(4.75	.46	5.00	1.32)	(4.86	.99	5.33	1.63)	
Educ. Father*	(5.00	.58	5.08	1.38)	(5.09	1.22	5.24	1.03)	
No. of Sibs	4.76	1.99	4.57	2.21	4.59	2.47	5.00	2.36	X
Income Level (1 = low, 2 = not low)	1.15	.36	1.29	.46	1.33	.48	1.36	.49	X
Ethnic Group (1 = White, 2 = Non-White)	2.00	00	1.91	.29	2.00	00	1.91	.29	
Age of Student (years)	11.2		11.2		12.1		12.1		

*Based on less than half the parents.

The fifth grade data are shown in Table 6.16. Nearly all measures (except of course WRAT Arithmetic) favored FT students. The difference on WRAT Reading is very impressive (more than one-half standard deviation). The other differences are on the order of one-fourth standard deviation. The changes in performance from third to fifth grade are as follows:

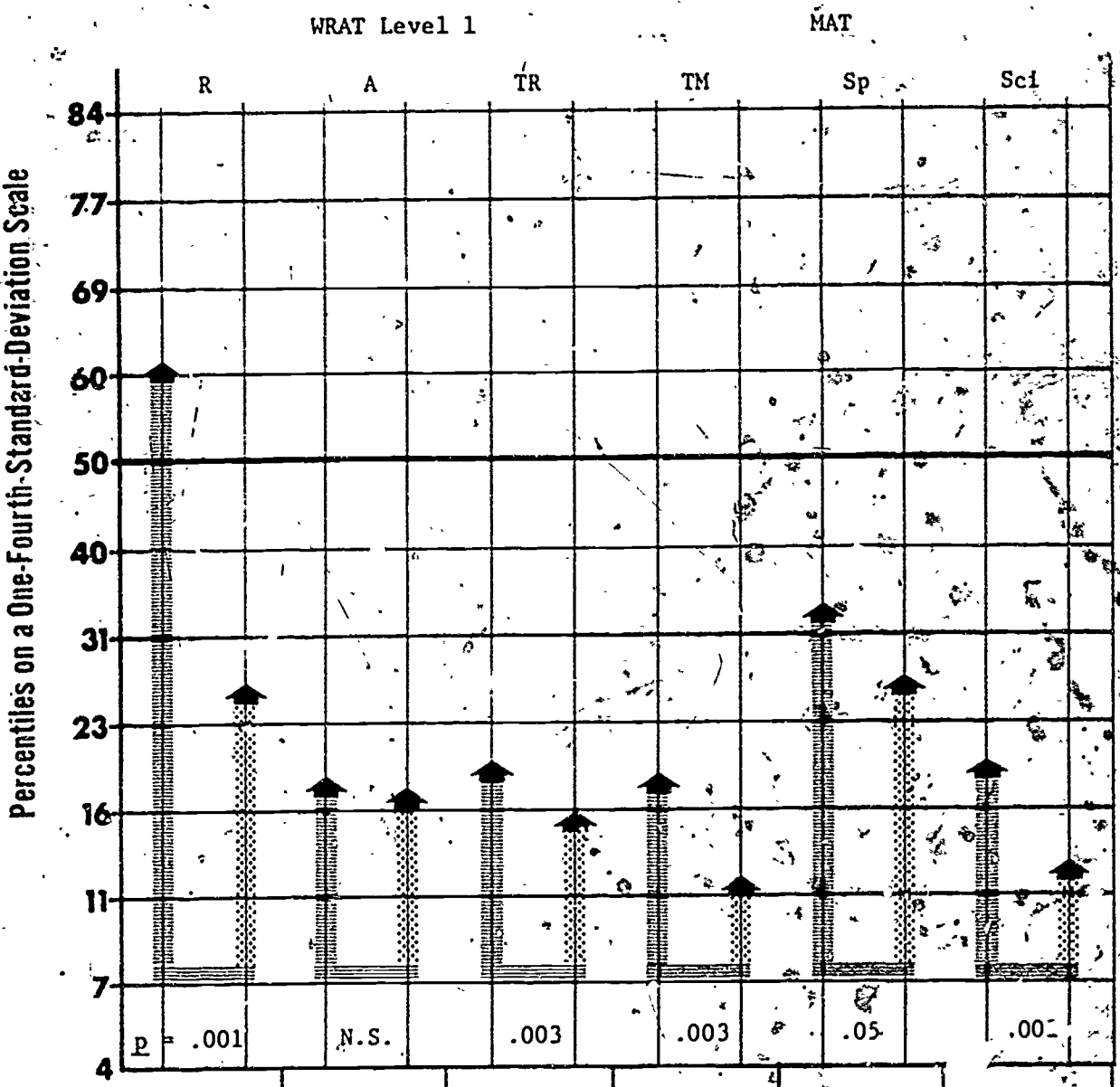
	Third			Fifth			Two-Year Gain (Loss)	
	%-tile	G. E.	N	%-tile	G. E.	N	%-tile	G. E.
WRAT								
Reading	69th	4.7	102	61st	(7.2)	93	-6	2.5
Arithmetic	50th	4.0	102	19th	(4.2)	95	-31	.2
MAT								
Total Reading	40th	3.3	102	20th	4.4	104	-20	1.4
Total Math	55th	4.0	102	19th	4.5	102	-36	.7
Spelling	40th	3.4	92	34th	5.0	104	-6	1.6

Table 6.16

No IQ data are available for the Dayton fifth and sixth graders. However, the fifth graders left third grade with a mean SIT IQ of 101.0 (N = 103) and the sixth graders with a mean of 112 (N = 33). The comparisons from third to fifth grade on the WRAT and MAT imply that spelling skills are being taught and reading decoding skills are being maintained. However, there are definite losses against the norm group on reading comprehension skills and math. In spite of these backward moves, the FT children are superior to the NFT children on 13 out of 14 measures.

Table 6.16

Dayton Fifth Grade FT-NFT Comparisons for Unadjusted Test Scores; p Values are for Covariance Adjusted Scores



GROUP	FT	NFT	FT	NFT	FT	NFT	FT	NFT	FT	NFT	FT	NFT
POST	5	5	5	5	5	5	5	5	5	5	5	5
N	93	55	95	59	104	87	102	83	104	85	102	85
%-tile	(61)	(27)	(19	(18)	20	16	19	12	34	27	20	13
S. S.*	71.8	63.6	31.2	31.0	68.0	64.9	77.3	74.6	74.0	71.4	73.2	69.5
S. D.*	14.6	14.1	3.5	3.5	10.3	10.0	9.0	9.1	10.2	10.5	9.4	7.5
G. E.	(7.2)	(5.6)	(4.2)	(4.2)	4.4	4.0	4.5	4.3	5.0	4.7	4.5	3.9

*WRAT scores are raw scores.

The sixth grade comparisons show a number of trends favoring FT students such as WRAT Reading, Math Problem Solving, and Science. However, no differences are significant after covariance adjustments. If full data on parental education were available, the result might be different. At the end of third grade, these cohort 2 children were at the 69th percentile on WRAT Reading (N = 33). On WRAT Arithmetic, they were at the 55th percentile. There are sharp losses on both WRAT Arithmetic and MAT Math against the norm group. Spelling is also very deficient for the sixth grade Follow Through group. This was not the case at the end of third grade (WRAT Spelling 73rd percentile versus 13th percentile on the MAT at sixth grade). A strong possibility is that these children are not being adequately taught in the upper elementary grades. These results are also perplexing in view of the mean IQ of 112 for these students at the end of third grade.

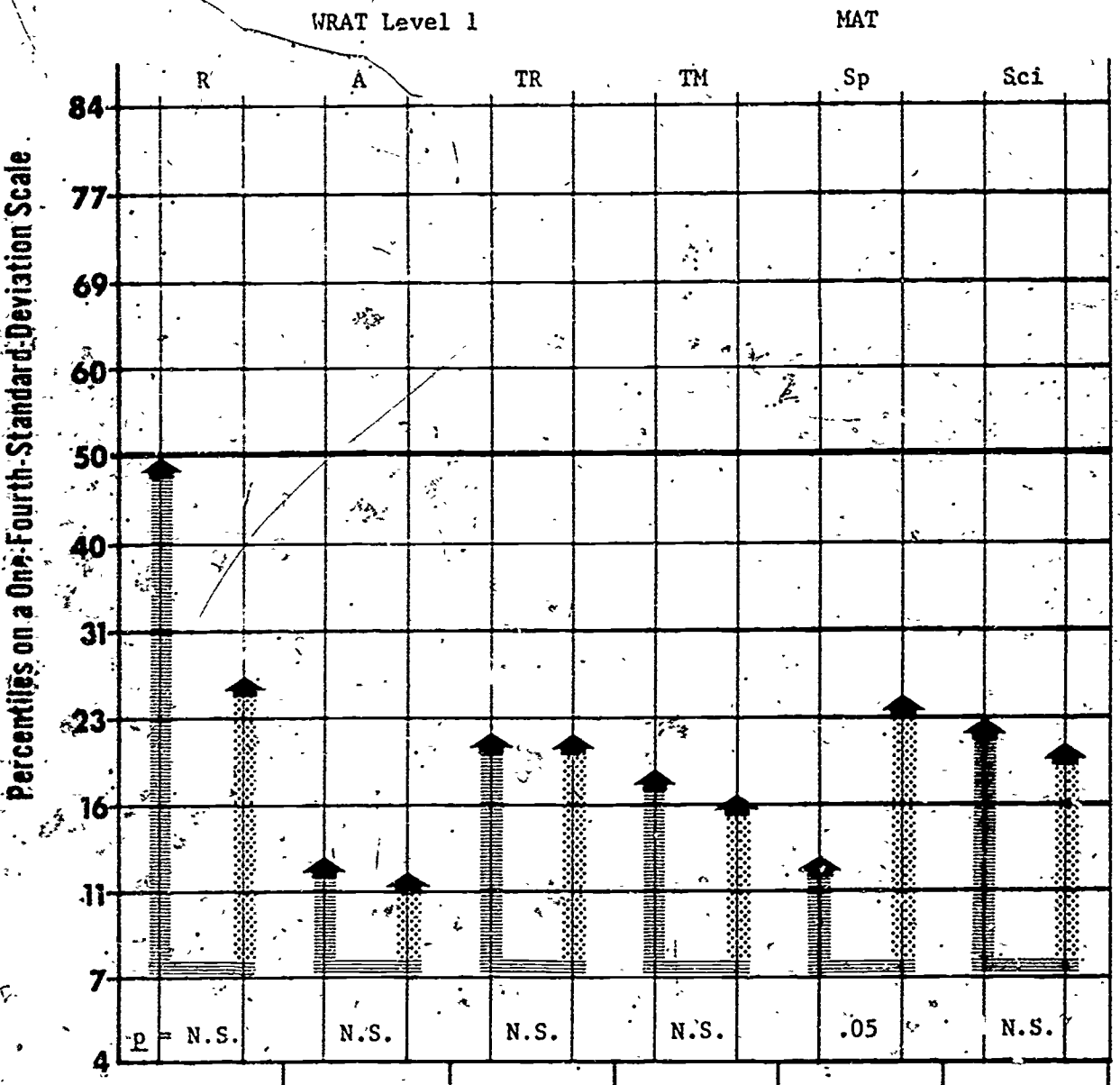
Table 6.17

Tupelo

The comparison children in Tupelo came from a variety of city and county schools. The background variables are summarized in Table 6.18. The control group has more Whites than the FT groups which are mostly Blacks. There appears to be a lower educational level for the sixth grade NFT parents and the fifth grade FT parents. However, these variables could not be used in the covariance analysis because of the large amount of missing data. The fifth grade NFT group also has more boys. The covariance adjustments that were made do not materially effect the group differences except on fifth grade MAT Total Reading, where the adjustment leads to a difference trend favoring Follow Through.

Table 6.17

Dayton Sixth Grade FT-NFT Comparison for Unadjusted Test Scores; p Values are for Covariance-Adjusted Scores



GROUP	FT	NFT	FT	NFT	FT	NFT	FT	NFT	FT	NFT	FT	NFT
POST	6	6	6	6	6	6	6	6	6	6	6	6
N	18	26	20	25	33	47	32	45	32	45	33	45
%-tile	(50)	(27)	(13)	(12)	22	22	19	17	13	25	23	21
S. S.*	75.1	69.4	32.6	32.7	74.3	73.9	83.5	82.5	71.6	77.4	78.4	77.3
S. D.*	14.4	15.2	2.9	4.0	7.8	14.0	8.1	10.2	11.0	11.8	8.2	9.9
G. E.	(7.0)	(6.0)	(4.6)	(4.6)	5.1	5.2	5.2	5.2	4.7	5.5	5.2	5.0

*WRAT scores are raw scores.

Tables 6.19 and 6.20 show the Follow Through children do significantly better on WRAT Reading and a trend for a better performance on California Achievement Test Reading (converted to MAT equivalents). FT children also are better on WRAT Arithmetic at the sixth grade, but we do not trust this test. The mean SIT IQ's of the fifth graders are a little lower than their comparison groups (88.9 versus 92.3). The mean IQ of these cohort 3 children at the end of third grade was 94.0 ($\underline{N}$ = 42). The FT sixth graders have significantly higher IQ's than their comparison group (92.5 to 84.4). This group left third grade with a mean IQ of 100.9 ($\underline{N}$ = 50). Again, it is difficult to interpret this difference, but in our judgment the NFT comparison group for sixth grade is a lower performing group. This conclusion is supported by the lower education levels of the parents on whom we have data. Thus, we would tend to discount the sixth grade data, but not the fifth grade data.

- - - - - - - - - - - - - - -

Tables 6.18, 6.19 and 6.20

- - - - - - - - - - - - - - -

Table 6.18

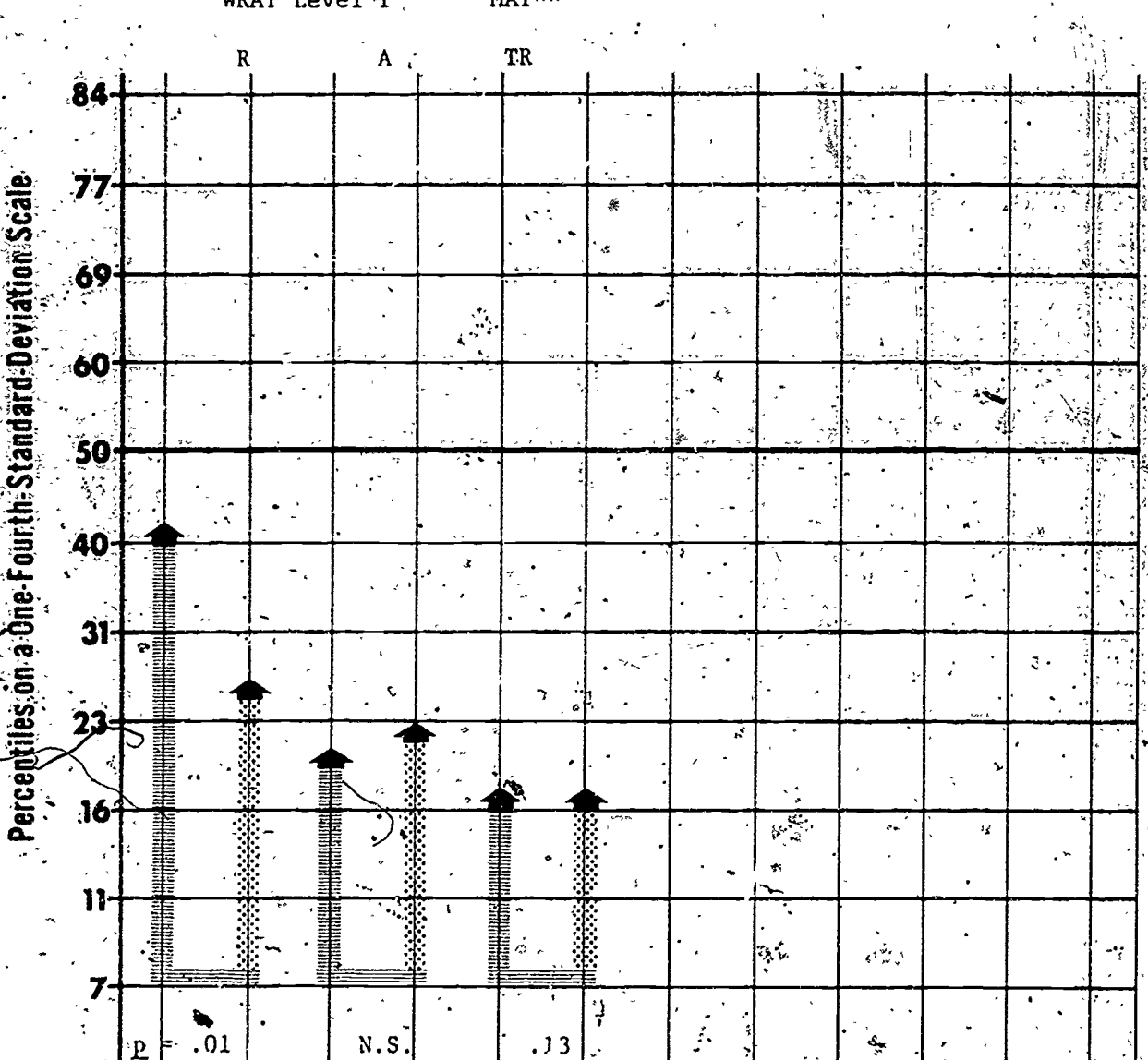
Background Variables for Tupelo FT-NFT Comparisons for Fifth and Sixth Grades

	Fifth				Sixth				
	FT		NFT		FT		NFT		
N	46		34		38		25		
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Used as Covariate (X)
Sex (1 = Male, 2 = Female)	1.48	.50	1.35	.48	1.55	.50	1.48	.51	X
Educ. Mother* (4 = High School Grad.)	(3.22	1.29	3.58	1.62)	(3.97	1.55	2.20	1.30)	
Educ. Father*	(3.56	1.00	4.12	1.15)	(3.89	1.17	2.60	.55)	
No. of Sibs	4.41	2.22	5.12	2.29	4.53	2.06	4.60	1.52	X
Income Level (1 = low, 2 = not low)	1.05	.23	1.03	.17	1.10	.31	1.00	00	X
Ethnic Group (1 = White, 2 = Non-White)	1.93	.26	1.53	.50	1.91	.29	1.68	.47	X
Age of Student (years)	11.2		11.2		12.0		12.4		

*Based on less than half the parents.

Table 6.19

Tupelo Fifth Grade FT-NFT Comparison for Unadjusted Test Scores; p Values are for Covariance Adjusted Scores

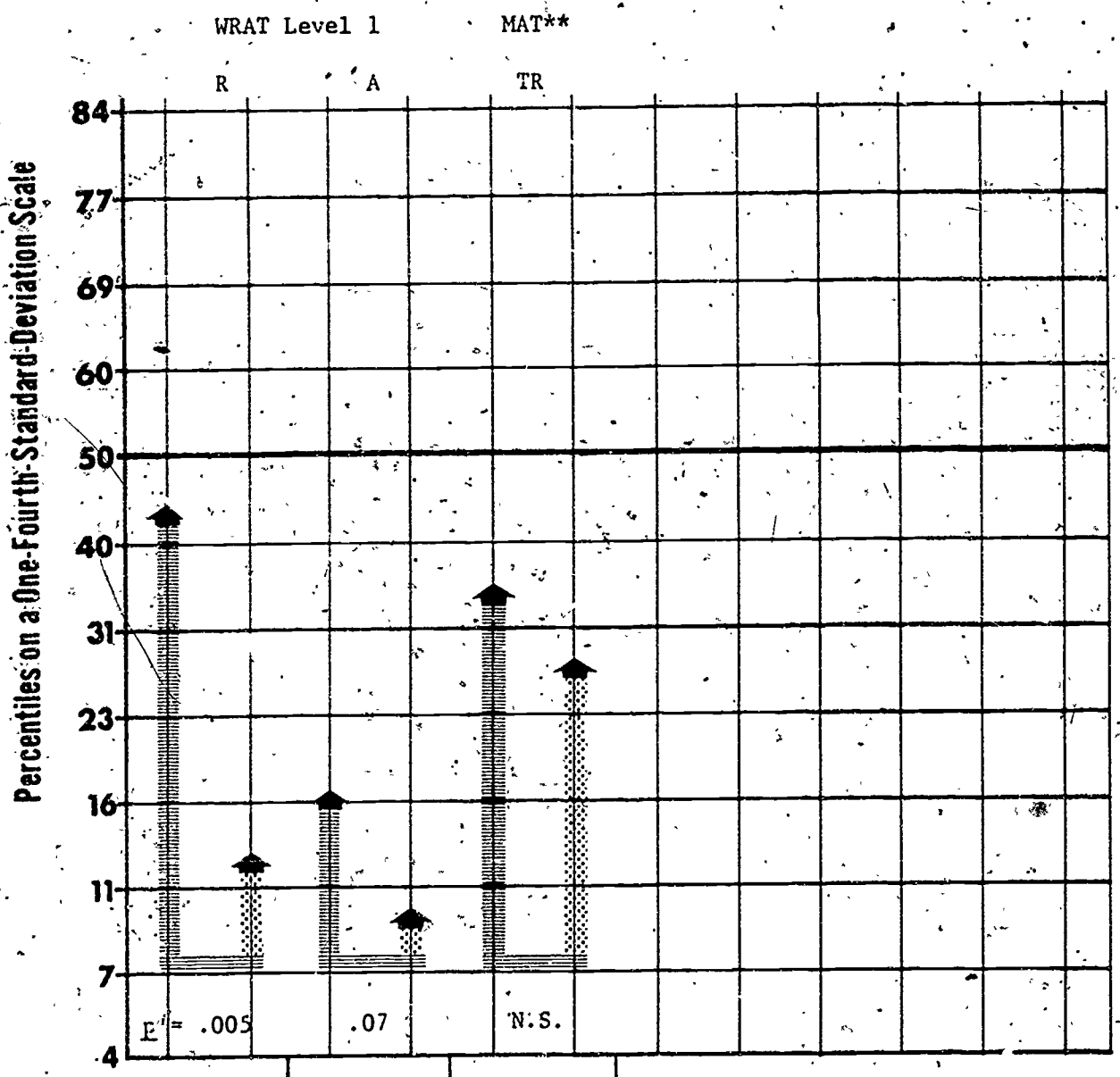


GROUP	FT	NFT	FT	NFT	FT	NFT
POST	5	5	5	5	5	5
N	56	34	56	34	56	33
%-tile	(42)	(27)	(21)	(23)	18	18
S. S*	67.3	63.3	31.5	32.5	66.8	67.0
S. D.	12.3	11.9	2.8	3.3	10.6	11.4
G. E.	(5.6)	(4.9)	(4.4)	(4.6)	4.3	4.3

*WRAT scores are raw scores.
**CAT converted to MAT equivalents.

Table 6.20

Tupelo Sixth Grade FT-NFT Comparison for Unadjusted Test Scores;
p Values are for Covariance Adjusted Scores



GROUP	FT	NFT	FT	NFT	FT	NFT
POST	6	6	6	6	6	6
N	38	25	38	25	37	24
%-tile	(44)	(13)	(17)	(10)	36	28
S. S. *	72.4	64.2	33.7	32.6	73.8	71.2
S. D.	9.8	10.6	4.2	3.1	9.7	9.0
G. E.	(6.6)	(5.0)	(4.9)	(4.6)	(5.0)	(4.6)

*WRAT scores are raw scores.
**CAT converted to MAT equivalents.

The Follow Through children showed changes in norm group standing from third grade as follows:

	Third			Fifth			Two-Year Gain	
	%-tile	G. E.	N	%-tile	G. E.	N	%-tile	G. E.
WRAT								
Reading	66th	4.4	42	(42nd)	(5.6)	56	- 24	1.2
Arithmetic	53rd	3.9	42	(21st)	(4.4)	56	- 32	.5
MAT								
Total Reading	28th	2.9	42	18th	4.3	56	-10	1.4
WRAT								
Reading	68th	4.4		(44th)	(6.6)	38	- 24	2.2
Arithmetic	46th	3.6		(17th)	(4.9)	38	- 29	1.3

While the WRAT Arithmetic data should not be taken literally, the general trend of the data support a loss against the norm group after leaving Follow Through. Reading decoding skills are maintained, however.

Flint

The comparisons for Flint are limited to three MAT Reading measures. They show no significant differences. The background variables presented in Table 6.21 show that information on parent education is lacking and that more than half of the comparison children are Non-Low Income, whereas only one FT student is. In our judgment these comparisons are not valid for judging the program in Flint. There are further complications due to the fact that most Title 1 children in Flint were in Distar programs from the beginning of Follow Through.

Looking at the progress of the K-starting cohort 2 children from the end of third grade on MAT Total Reading, we see a drop from the 34th percentile (N = 41) to the 22nd percentile (N = 35). The two-year gain in grade equivalents was from 3.14 to 4.35.

Table 6.21

Background Variables for Flint FT-NFT Comparisons for Fifth Grade

	Fifth			
	FT		NFT	
N	35		82	
	Mean	SD	Mean	SD
Sex (1 = Male, 2 = Female)	1.51	.51	1.57	.55
No. of Sibs	4.11	1.59	4.65	2.18
Income Level (1 = low, 2 = not low)	1.06	.24	1.57	.55
Ethnic Group (1 = White, 2 = Non-White)	1.94	.24	1.77	.42
Age of Student (years)	11.0		11.2	

Todd County

Although not a part of the controlled comparisons, a sample of Todd County fifth and sixth graders was also tested. Since we have third grade MAT data only on the fifth graders, we will present only that data so that gains from third grade can be examined.

	Third			Fifth			Two-Year Gain (Loss)	
	%-tile	G. E.	N	%-tile	G. E.	N	%-tile	G. E.
Total Reading	42nd	3.4	43	27th	4.7	44	- 15	1.7
Total Math	44th	3.8	43	20th	4.7	44	- 24	1.1
Spelling	52nd	3.9	43	38th	5.2	44	- 14	1.4

The results show some loss against the norm group, but not as drastic as for other places. Some reasonable gains are made on MAT Reading after Follow Through and on Spelling.

Summary

The follow-up study of more than 700 fifth and sixth graders, along with comparison groups, clearly showed measurable persisting effects of the largely 3-year E-B Follow Through programs.* The significant effects were particularly strong in reading on both the MAT and the WRAT. In mathematics, the MAT showed a number of significant effects, especially for Problem Solving. Significant Follow Through effects were also found on MAT Science.

Examination of the equality of the groups on background variables and the magnitude of effects by site revealed several important points. A few of the comparisons appear to be invalid because the Follow Through and comparison groups were not adequately matched. However, the overall effects are not changed by this. It appears that some spelling is taught after third grade which helps to keep percentiles on spelling close to where they were at the end of Follow Through. Reading decoding skills also tend to be maintained. However, large losses against the norm group are apparent in math for most sites. The implication is that effective teaching of math at the upper elementary levels in our Follow Through sites is not occurring. Status on the MAT reading comprehension measure at the end of fifth grade is typically low in the Non-White communities (Tupelo, Uvalde, Dayton, E. St. Louis) for the cohorts studied. In several cases, this measure was also low at the end of third grade; in others there have been sizeable losses against the norm group. We expect future cohorts to do better on this measure, especially those starting the program in kindergarten.

*In 1976 this study was repeated with 600 Follow Through fifth and sixth graders and 600 comparison students in six sites. The result was nearly identical.

To our way of thinking, the data support once again the conclusion drawn from Head Start data that gave birth to Follow Through. Gains made through better instruction in Head Start were lost when the children were put through into an educational system that was not effective. There is clearly an implication and a need to study the effects of continuing a Follow Through type program until adult competencies are mastered.

In our view the basic deficiency in performance of many children of poverty is in the size of their vocabulary (language concepts). To a large extent, language concepts are taught at home, not school, for middle-class kids. By the end of third grade, the typical home and school have taught an oral vocabulary of from 5000 to 8000 words, and a reading vocabulary of 2000 words. The children are then thrust upon testbooks (and tests of reading comprehension) where comprehension of the full adult vocabulary is expected. (Thorndike-Lorge estimate that the typical high school senior has a reading vocabulary of 15,000 concepts). It is not surprising that children of poverty are more likely to flounder at this time. As we see it, the E-B programs have overcome much of the gap in language comprehension, but in 1st-starting sites it might take two more years of extensive language instruction to "lick the problem," and in K-starting sites another year. An alternative solution is to change educational systems so that there is systematic language instruction in the schools (K to 12) which progressively builds an adult reading vocabulary of 15,000 words year-by-year, and does not leave the instruction primarily up to parents, TV, and whatever. We are developing approaches to dealing with this problem in the near future.

CHAPTER 7

COMPARISON OF PROGRESS IN RELATION TO IQ

The Low IQ Study

The data for all children in our main file with IQ's on the Slosson Intelligence Test under 80 at entry to the program were compared with that for our total group (All). The data were analyzed separately for Pre-K to Post-K and Post-K to Post-3, and for Pre-1 to Post-1 and Post-1 to Post-3. The data show that Low-IQ children (mean 73 at pretest and ranging as low as 50) gain more than a year on WRAT Reading for each year of instruction. On the average, the gain is approximately 1.2 grade levels each year, while the gain for the All group averages 1.35 grade levels each year.

- - - - - -

Figure 7.1

- - - - - -

The gains on WRAT Arithmetic are shown in Figure 7.2. The results are very similar. K-starting Low IQ students actually make larger gains in kindergarten than the All group and show average gains over four years of .95 grade equivalents. The average gain over four years for the All group is 1.00. The 1st-starting Low IQ students nearly match the gains of All group during the first year and over three years. The average gain for the Low IQ group is 1.04 grade equivalents and for the All group 1.07.

- - - - - -

Figure 7.2

- - - - - -

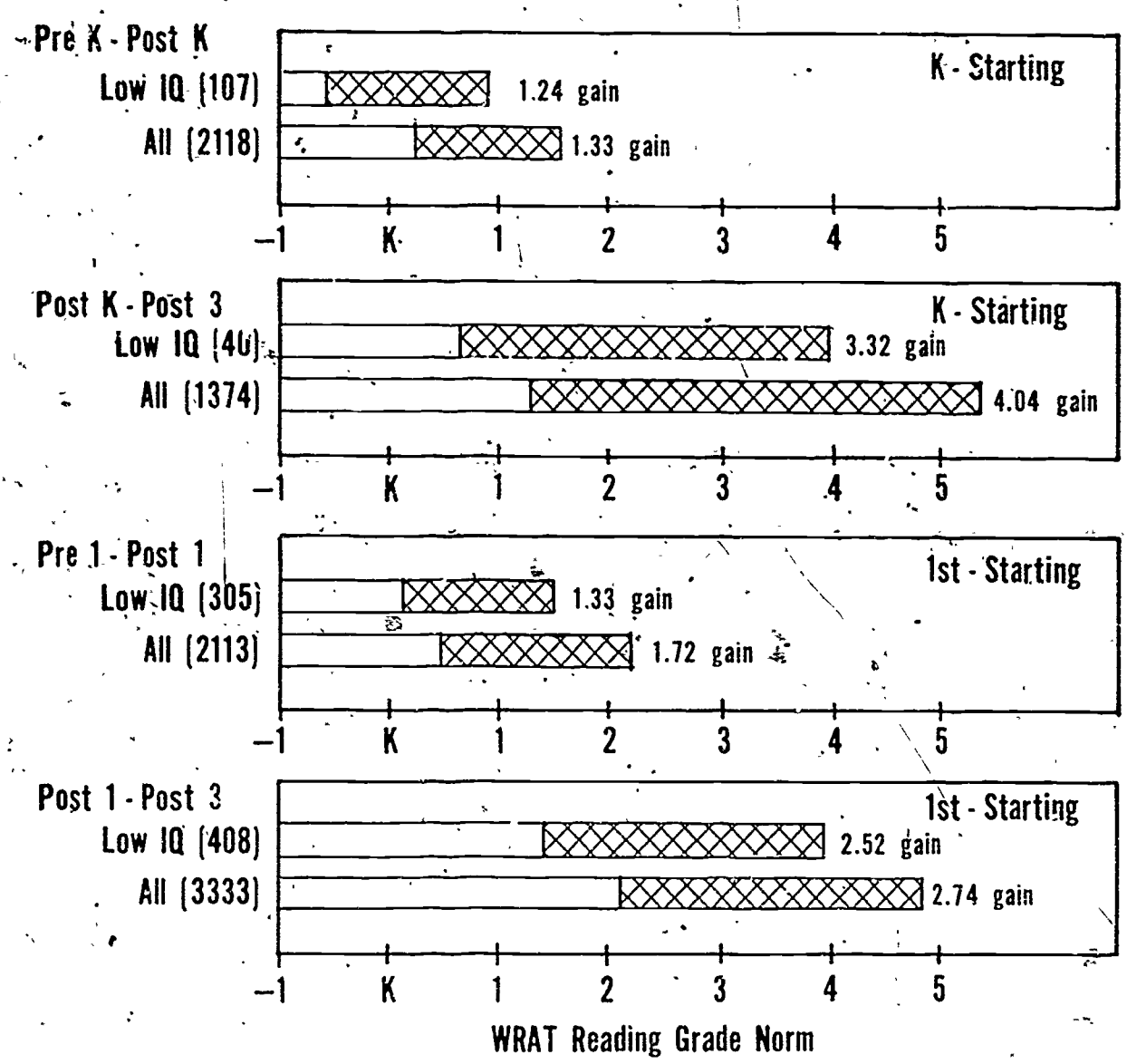


Figure 7.1 Reading gains of low IQ disadvantaged children. (Shaded area indicates gain for the time period shown to the left of each chart.)

Low IQ = IQ of 80 or less in Engelmann-Becker Follow Through Program.
All = All children in Engelmann-Becker Follow Through Program.

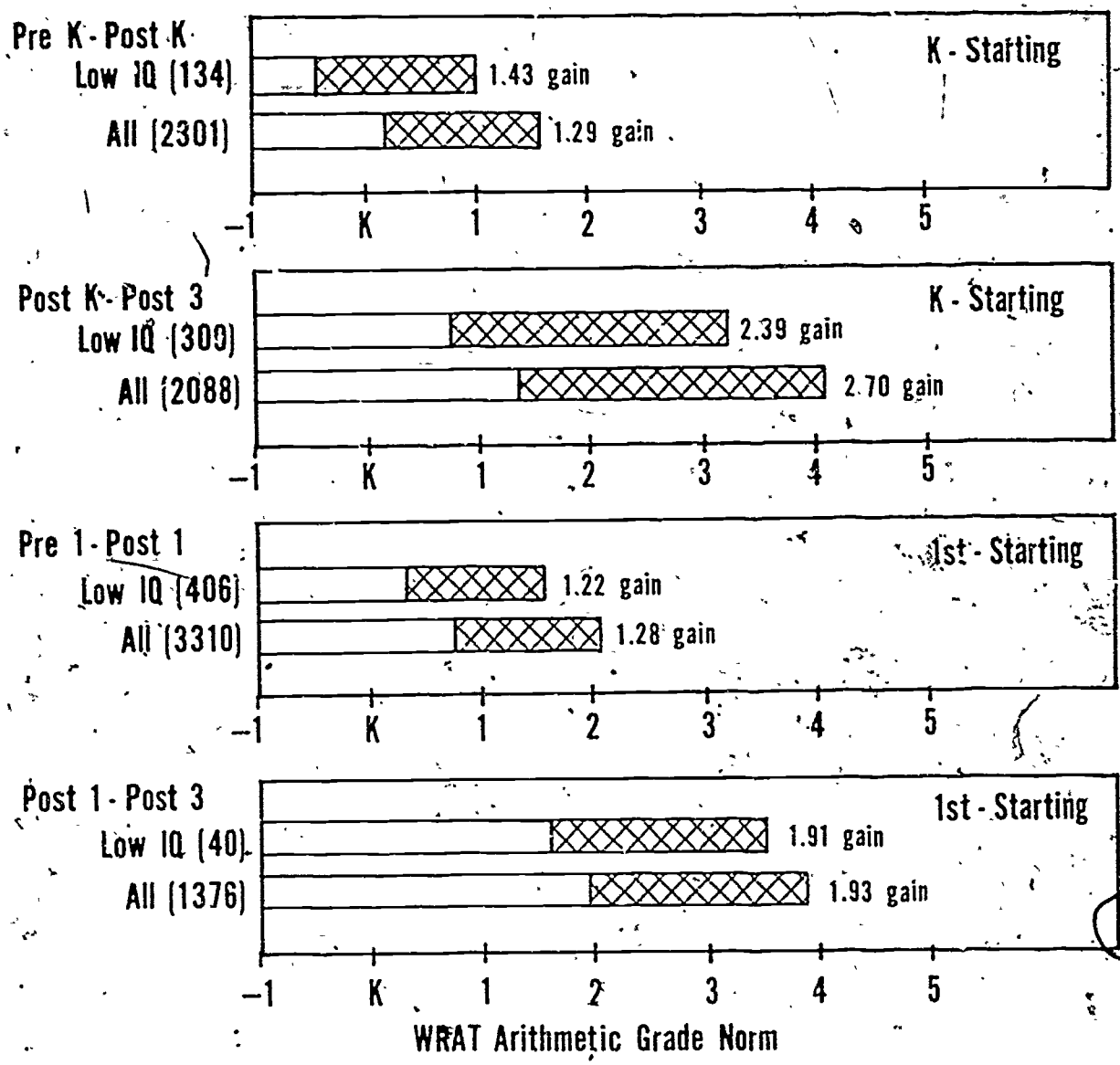


Figure 7.2. Arithmetic gains of low IQ disadvantaged children. (Shaded area indicates gain for the time period shown to the left of each chart.)

Low IQ = IQ of 80 or less, in Engelmann-Becker Follow Through Model.
All = All children in Engelmann-Becker Follow Through Model.

Table 7.1

Mean Gains in Standard Scores for WRAT Reading and Arithmetic

	Pre K - Post K		Post K - Post 3rd	
Reading	Mean Gain	SD Gain	Mean Gain	SD Gain
Low IQ	24.9	20.6	11.2	17.7
ALL	24.5	18.0	8.5	18.0
Arithmetic				
Low IQ	27.7	15.8	.72	10.0
ALL	23.6	16.2	-5.2	11.6

	Pre 1st - Post 1st		Post 1st - Post 3rd	
Reading	Mean Gain	SD Gain	Mean Gain	SD Gain
Low IQ	14.5	13.3	9.0	6.9
ALL	19.4	15.8	12.3	16.0
Arithmetic				
Low IQ	11.6	11.6	2.3	8.7
ALL	11.0	11.3	0.0	9.8

The mean gains in standard scores are given in Table 7.1. Except for the 1st-starting group in reading, the gains for the Low IQ Group are equal to or better than those for the ALL Group.

An examination of the number of lessons taught each year to the various groups shows that Low IQ children are taught 30 fewer Distar lessons in Kindergarten (160 lessons is average) and 25 fewer in first grade. After that the differential drops to less than 15 lessons. There is an implication that what a child can learn is less a function of IQ and more a function of method of instruction.

Table 7.1 shows the IQ gains by groups used in the analysis. Statistical regression effects would make the Pre-K to Post-K and Pre-1 to Post-1 gains for the Low IQ groups greater than the true gains. The average regression can be predicted by multiplying the deviation of the Low IQ group from 100 by .92 (reliability of SIT IQ). For the K-starting group whose mean is 73, this procedure gives an estimated true deviation score of (73 - 100) x .92 = -24.8 or an estimated true mean IQ of 75.2. The K-starting and 1st-starting ALL groups both show net IQ gains Pre-K to Post-3 of about 8 points. No regression effects influence these gains. Thus, we would estimate the true gains for the Low IQ students to be somewhere between 8 and 14 points.

Table 7.2

Table 7.2

Slosson IQ GAINS for Low IQ and ALL Groups

Pre K - Post K	Pre		Post	Gain	Adjusted Gain	SD Gain	N
Low IQ	73.0	(75.2)*	93.8	20.8	(17.6)**	12.1	142
ALL	103.3		112.8	9.5		14.2	2270
Post K - Post 3	**Pre**		**Post**	**Gain**	**Adjusted Gain**	**SD Gain**	**N**
Low IQ	80.3	(81.8)*	87.8	7.5	(7.0)**	15.0	38
ALL	109.3		107.9	-1.4		14.0	1305
Pre 1 - Post 1	**Pre**		**Post**	**Gain**	**Adjusted Gain**	**SD Gain**	**N**
Low IQ	72.5	(74.7)*	85.5	13.0	(10.8)**	10.8	319
ALL	94.7		100.1	5.5		12.0	2113
Post 1 - Post 3	**Pre**		**Post**	**Gain**	**Adjusted Gain**	**SD Gain**	**N**
Low IQ	78.7	(80.4)*	83.4	4.7	(3.0)**	12.9	387
ALL	99.2		101.6	2.5		13.4	3081

* Estimated true score means.

** Estimated true gain.

Relation to Entry IQ to Academic Performance

An analysis was performed to see if children with different entry SIT IQ's benefit more or less from the program. The general result is illustrated in Figures 7.3 and 7.4. Level of performance in standard scores or grade equivalents shows a systematic increase with higher starting IQ's. For the most part, however, the year-to-year gains do not systematically increase as starting IQ increases. Children with higher entry IQ's start higher on most achievement measures (reading, arithmetic, spelling) and stay higher, but do not usually gain proportionately more.

Figures 7.3 and 7.4 about here

There are two exceptions to this conclusion which are illustrated in Table 7.3 and 7.4. In going from the Primary II MAT Total Reading (end of second grade) to the Elementary Level (end of third grade), a clear relationship between amount of gain and IQ blocks is apparent. This finding would be expected if general language development (which is fostered to a large extent outside of school) plays a key role in both the IQ measure and the MAT Elementary Reading Test. This finding is consistent with our interpretations of other findings on the MAT Elementary Reading Test, and will be treated in more detail in the discussion. The other exception is the first-year gain (Kindergarten for K-Starting and First Grade for First-Starting) on WRAT Reading. Children with higher entry IQ's gain more in reading decoding skills in their first year of instruction. This may be because their better language skills give them an advantage in beginning instruction, (e.g. some children do not have to learn the language of instruction before learning to decode).

Figure 7.3

WRAT Reading Mean Standard Scores for 1st-Starting Children by Entry SIT IQ Blocks

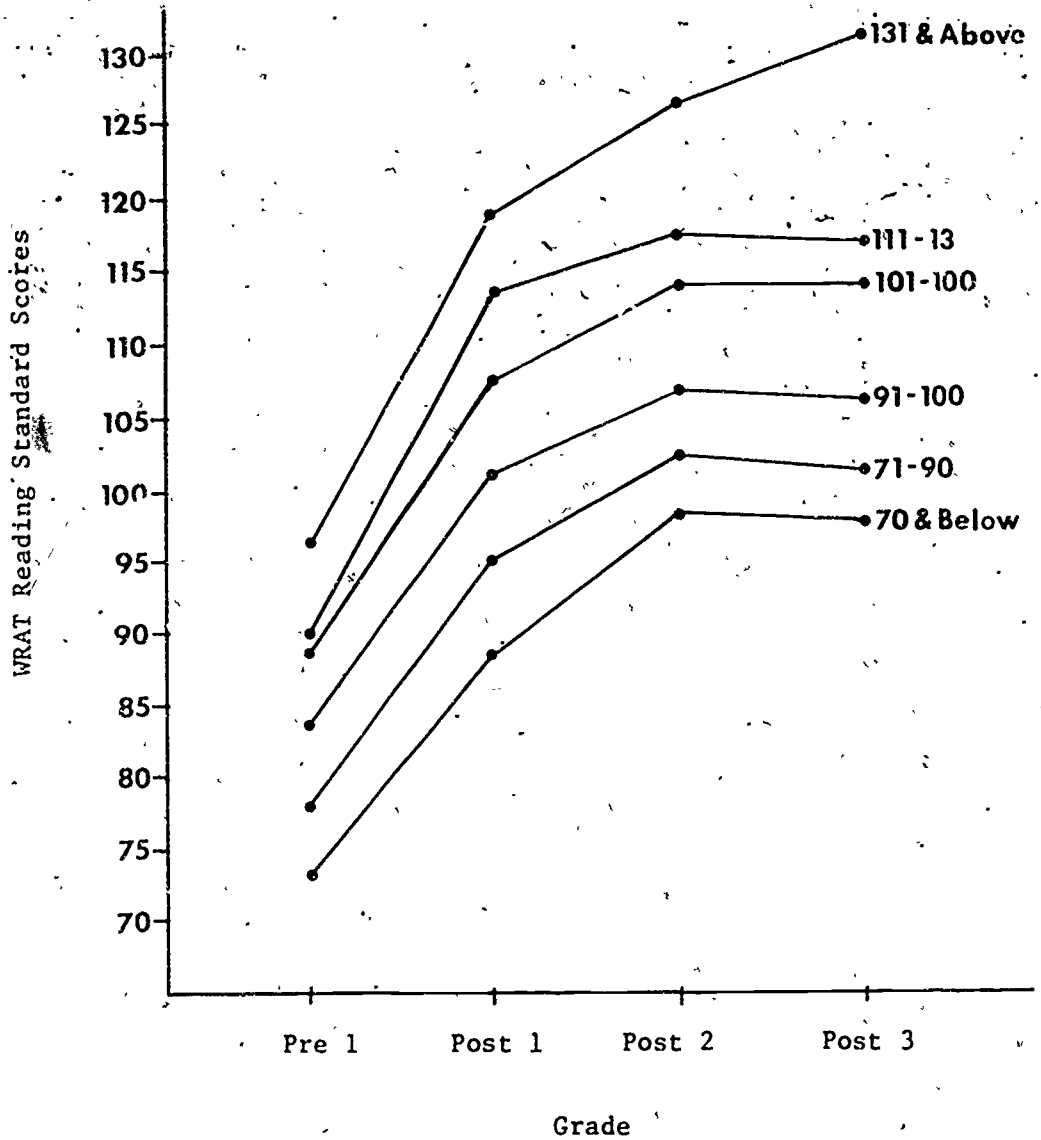


Figure 7.4

MAT Total Reading Mean Grade Equivalent Scores
for 1st-Starting Children by Entry SIT IQ Blocks

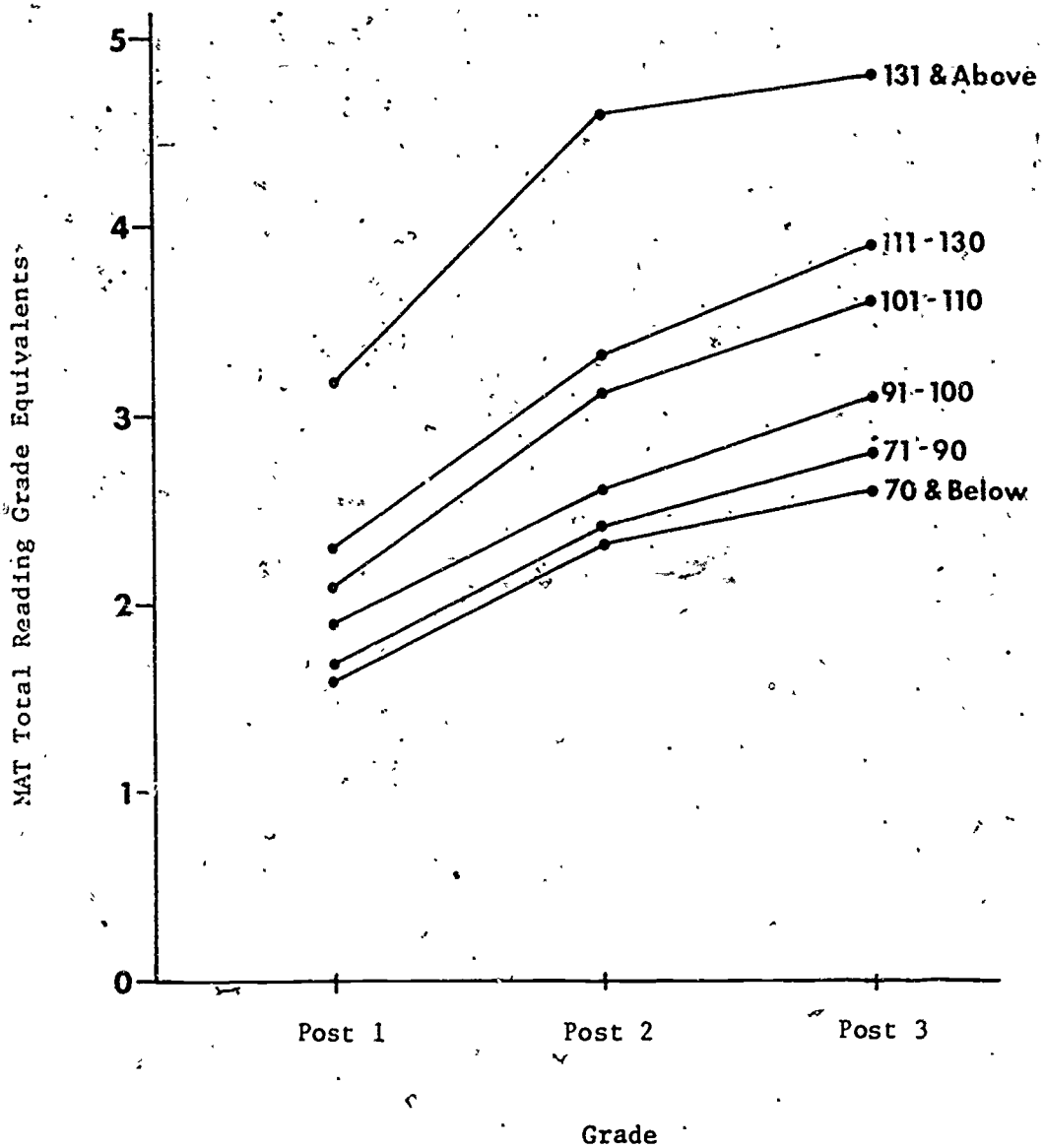


Table 7.3

WRAT Reading Standard Score Gains by Grade Level for 1st-Starting Students by SIT IQ Blocks

IQ Block	First Grade Mean Gain	First Grade N	Second Grade Mean Gain	Second Grade N	Third Grade Mean Gain	Third Grade N
70 and below	15.1	49	9.4	53	-2.5	42
71-90	16.9	366	8.1	428	-1.4	400
91-100	19.8	328	5.2	373	-1.2	353
101-110	21.1	214	5.3	241	.4	225
111-130	25.3	161	3.9	193	1.1	167
131 and above	24.0	25	8.8	28	5.9	17
Average	19.8	1143	6.2	1316	- .62	1204

MAT Total Reading Standard Score Gains by Grade Level for 1st-Starting Students by SIT IQ Blocks

IQ Block	Second Grade Mean Gain	Second Grade N	Third Grade Mean Gain	Third Grade N
70 and below	13.8	14	3.3	23
71-90	14.2	122	4.1	192
91-100	13.3	106	4.6	198
101-110	13.9	77	6.3	107
111-130	14.5	44	6.8	58
131 and above	12.4	14	—	(4)
Average	13.8	377	4.9	582

Table 7.4

WRAT Reading Standard Score Gains by Grade Level for K-Starting Students by SIT IQ Blocks

IQ Blocks	Kindergarten Mean Gain	N	First Grade Mean Gain	N	Second Grade Mean Gain	N	Third Grade Mean Gain	N
71-90	19.5	18	12.9	39	-2.6	43	-3.5	42
91-100	15.8	52	9.3	99	-3.7	100	-.39	91
101-110	19.6	64	11.4	124	-4.2	133	2.86	123
111-130	24.2	61	9.3	142	-6.1	149	1.	138
131 and above	29.2	12	7.4	22	-1.5	22	-2.52	23
Average	20.5	209	10.1	429	-4.4	450	1.36	420

MAT Total Reading Standard Score Gains by Grade Level for K-Starting Students by SIT IQ Blocks

IQ Blocks	Second Grade Mean Gain	N	Third Grade Mean Gain	N
71-90	9.0	15	2.3	41
91-100	9.7	37	3.7	82
101-110	7.6	49	5.0	115
111-130	8.7	76	6.8	128
131 and above	9.4	14	6.1	23
Average	8.7	192	5.1	392

Summary

This evidence demonstrates that the Direct Instruction Model is clearly effective with Low IQ children. This is the group who have been hardest to teach in the past. With good instruction that builds on their entry skills in a systematic way, Low IQ children can be taught to be smarter and more capable.

The evidence also demonstrates that entry IQ is not a major determinant of academic gains in the direct-instruction programs. Where some relationships to entry IQ's are found, they seem closely tied to the role of general language competency in what is being measured (Elementary MAT) or to the language skills necessary to facilitate beginning instruction.

CHAPTER 8

FORMATIVE EVALUATION OF DIRECT INSTRUCTION

Douglas W. Carnine

Direct Instruction's effectiveness, as indicated by the earlier chapters in this report, has also been documented in other Follow Through research [1, 2, 3] and in summaries of Distar research conducted in non-Follow Through settings.[4, 5] Although the data indicate that direct instruction is effective, questions concerning direct instruction's functional components remain. The Direct Instruction Program's components include teaching techniques, procedures for teacher training, and the Distar curriculum, to name a few. Identifying the functional components are important for both developing a technology of education and disseminating the program. A technology of education requires that each teaching technique, each training procedure, and each curriculum design procedure be evaluated. Identifying what works is critical in dissemination because of the limited resources available--school districts cannot afford to purchase ineffective curricula or weeks and weeks of consultant time. Only by identifying the techniques and procedures that work and discarding the others can a science of teaching and an affordable dissemination plan evolve.

This chapter's purpose is to summarize our research concerning the teaching techniques, training procedures, and curriculum design procedures that comprise the Direct Instruction Program. The research on teaching techniques and training, which will be discussed first, addresses three questions. To what extent do the teaching techniques effect child learning? Can inexperienced teachers implement the techniques as well as the experienced teachers? And finally is training necessary for inexperienced teachers to learn the techniques? The findings concerning curriculum design procedures will follow the findings on teaching techniques and training.

Teaching Techniques. The teaching techniques that we have investigated are pacing, corrections, reinforcement, feedback, and signals. Rapid pacing (see Appendix B. 1) resulted in more correct responses and fewer off-task responses than slow pacing during Distar Reading I instruction. (When the teacher asked about 12 questions per minute, the children answered correctly about 80 percent of the time and were off task only about 10 percent of the time. When the teacher asked only four questions per minute, the children answered correctly about 30 percent of the time and were off task about 70 percent of the time.)

Corrections following errors (see Appendix B. 2) resulted in more correct training and posttest responses than no corrections during arithmetic instruction. (When the teacher corrected errors, the children answered correctly about 70 percent of the time during training and 65 percent of the time during posttest. When the teacher did not correct errors, the children answered correctly about 15 percent of the time during training and 15 percent of the time during posttest.) In

another study, (see Appendix B. 3) when children who made frequent errors decoding simple words were precorrected, they read more words correctly. (When the teacher precorrected by asking the child to identify the vowel sound before reading the word, the percent of words read correctly increased from 15 percent to 63 percent for one child and from 12 percent to 43 percent for the other child.)

In the reinforcement study (see Appendix B. 4), social praise for on-task behavior increased on-task behavior, for the children. (When the teacher praised on-task behavior, the children were on task over 80 percent of the time. When the teacher did not praise on-task behavior, the children were on task about 50 percent of the time.) In the feedback study (see Appendix B. 5), children made fewer errors on their arithmetic worksheets when they were told how many errors they made and graphed their performance. Since the research concerning reinforcement is extensive, it has received relatively little attention in our research program.

A signal is a cue used during small group instruction to indicate that the children are to respond in unison. In the signal's studies (see Appendix B. 6), teacher signals resulted in small increases in the children's attending and responding. (When the teacher used signals, the children attended about 55 percent of the time and responded about 80 percent of the time. When the teacher did not use signals, the children attended about 35 percent of the time and responded about 60 percent of the time.) Because the effects from using signals were relatively small and because data relating signals to academic responding is inconclusive, signals as a teaching technique is being emphasized less in our training. Current research on academic performance in a small group teaching

situation is comparing group responding and individual responding. To the extent that group responding enhances achievement and signals increase group responding, training teachers to use signals will be a worthwhile training activity.

Training teachers. Developing an empirically based teacher training program requires knowledge about which curricula and which techniques enhance children's learning. The outcome data on the Distar programs and the teaching techniques research provide that necessary curricular and technique knowledge. Unfortunately the research program on teaching techniques has only recently demonstrated the effectiveness of the various techniques. Since the teacher training research was dependent on these findings, the teacher training research is not as extensive as is the research on teaching techniques.

To answer our second question--whether inexperienced teachers can implement the techniques as well as experienced teachers--we observed 13 inexperienced Distar teachers twice weekly over a three-month period (see Appendix B. 7). We compared the inexperienced teachers' implementation as they received training--preservice, inservice, and in-classroom--with our experienced teachers' implementation as measured during the technique evaluation research.

In the corrections study cited earlier, the teacher corrected approximately 90 percent of the child errors. The inexperienced teachers were trained to correct about 60 percent of the child errors. Naive teachers can be trained to ask approximately seven questions per minute in contrast to the experimental teachers' rate of 12 per minute. Finally, the observations indicated that teachers can be

trained to use clear signals on approximately 75 percent of the tasks they present. Although the training procedure used in this observation study was not evaluated, the staff felt that the sequence for training the teachers simplified the training process and was more efficient than the procedures previously used in Follow Through.

The third question was whether preservice and inservice training were necessary for teachers to adequately pace their presentations and use signals (see Appendix B. 7). Using a multiple baseline design with three teachers, appropriate signaling increased from 29 percent to 92 percent for one teacher, from 26 percent to 98 percent for another, and from 29 percent to 96 percent for the third. Pacing (questions per minute) increased from 3.8 to 6.8 for one teacher, from 3.8 to 11.8 for another, and from 3.6 to 9.4 for the third.

Answering criticisms about Distar. Some studies have been conducted to evaluate criticisms about the use of Distar in Follow Through. One criticism was that Distar was inappropriate for the higher performing children in Follow Through classrooms. In one study (see Appendix B. 8) a classroom of high performing children who had received two years of Distar reading was evaluated in terms of achievement and attitude. The children scored almost two years above grade level in reading as measured by the Stanford Achievement Test and the questionnaire responses indicated that the children enjoyed the program, thought of themselves as intelligent, and had been stimulated to pursue topics introduced in Distar during their free time. In a second study (see Appendix B. 9), middle-ability first graders and high-ability second graders were taught fractions by either the Distar arithmetic procedure or by a more

inductive (practice only) approach. The first graders taught according to tne Distar procedures scored higher on two different transfer tests--one consisting of the same fraction skills, but with new examples, and the other consisting of untaught fraction skills. The Distar second graders scored significantly higher on only the first transfer test.

The final two studies in this section investigated the criticism that Distar involves nothing but rote learning--the children don't understand and can't transfer what they learn to new problems. The first study focused on the extent to which Distar arithmetic teaches 'understanding'. Learning disabled first graders were randomly assigned to Distar Arithmetic I instruction for one year or to a school district's lab math program which stressed understanding and the use of manipulatives. The Distar students not only learned more arithmetic skills during the year but also did as well as the lab math children in solving story problems that involved manipulatives, even though the Distar children never received instruction in working story problems by using manipulatives. The Distar children out performed the lab math children on story problems that were presented verbally. The second study (see Appendix B. 10) focused on how well children can transfer the decoding skills they learn in early Distar Reading I to words that they have never before encountered. Preschoolers who were taught by a Distar approach correctly read 92 percent of the transfer words; preschoolers taught by a sight or whole-word approach correctly read 28 percent of the transfer words.

Curriculum Design Procedures. Although the research in this area is only indirectly related to Follow Through's purposes, the procedures under investigation were used in designing the Distar curricula and may, in part, explain the Engelmann-Becker Model's success. The programming

procedures are discussed in detail by Becker, Engelmann, and Thomas[6] in their book on cognitive learning. The research relates to five programming procedures: selecting positive and negative concept examples, modifying the examples to make them easier to learn, sequencing the examples, providing practice, and teaching component skills of complex operations before teaching the operation itself. The following suggestions that relate to the procedures are tentative and generalizations are restricted to the context of the actual studies.

Positive and negative examples should be selected so that students learn the essential characteristics that define the concepts. When the essential characteristic assumes a range of values (e.g., the concept red is defined by a range of color values), teachers should select positive and negative example pairs at each end of the range and some positive examples from within the range (see Appendix B. 11). Only intended essential characteristics should appear in positive examples. Characteristics that appear in all positive examples and are absent from negative examples may be treated as essential by the student, whether a teacher intends for them to be treated that way or not (see Appendix B. 12). Teachers should be careful that the relevant characteristics of the negative examples are not fewer or more obvious than the intended essential characteristics of the positive examples; otherwise a student may learn to respond to the presence of the negative characteristic rather than to the presence of the positive characteristic (see Appendix B. 13).

The procedures for modifying positive and negative examples are intended to make the essential characteristic(s) of the concept more salient to the student. Students tend to learn a concept's essential

characteristic more quickly if the characteristic is altered to make it more dissimilar from characteristics of other examples (see Appendix B. 14). Another procedure for increasing essential characteristic saliency, adding an emphasizer such as a color, tends to quicken learning when the concept is defined by a single essential characteristic. When a concept is defined by several essential characteristics, emphasizers may slow child acquisition (see Appendix B. 15).

The sequencing procedures are also intended to make essential characteristics more salient to the student. For concepts defined by several essential characteristics, presenting positive and negative example pairs that differ only in terms of a single essential characteristic results in more rapid acquisition than presenting example pairs that differ in terms of several characteristics; also pair presentations are more effective than successively presenting examples (see Appendix B. 16). The third sequencing procedure is intended to increase characteristic saliency by separating similar examples from each other in the order of introduction. The separation allows for students to discriminate positive examples from negative examples that are more dissimilar. Later the more similar examples would be introduced (see Appendix B. 17).

The practice procedure, called cumulative introduction, can lead to more rapid acquisition of a set of concepts. In cumulative introduction positive examples are introduced one at a time and are repeatedly presented until the students consistently identify each example introduced up to that point before new concept examples are introduced (see Appendix B. 17).

The final procedure is that component skills of complex operations are taught before the operation itself is introduced. For most students this procedure results in higher transfer scores to new problems, when it is compared with just teaching the operation without first teaching the component skills (see Appendix B. 9 and Appendix B. 10).

Although the research support for the selection, modification, sequencing, practice, and component skill teaching procedures is limited and has methodological flaws, it still indicates that the procedures used in designing the Distar curricula have some empirical support. Additional studies with greater methodological rigor are needed to evaluate these and other design procedures exemplified in Distar. A well-grounded empirical base for the procedures would help account for direct instruction's effectiveness and could be useful to other educators in designing instructional programs.

References and Notes

Chapter 1

1. Egbert, R. L. Planned Variation in Follow Through. Unpublished manuscript. Lincoln, Nebraska: College of Education, 1974.

2. For a fuller description of the Model, see the 1973-74 University of Oregon Follow Through Proposal; or write to the Follow Through Project, College of Education, University of Oregon, Eugene, Oregon, 97403, for the "Yellow Book"; or by the end of 1976 write to Oregon for the Program Implementation Manual (cost about $8.00).

3. Follow Through Technical Report, 1969-1971, page 3-B. Menlo Park, California: Stanford Research Institute, 1972.

4. Education as Experimentation: A Planned Variation Model. Volume II B, page C-2. Cambridge, Mass.: Abt. Associates Inc., 1975.

5. Education as Experimentation: A Planned Variation Model. Volume III. Four-Year Effects of Follow Through. Cambridge, Mass.: Abt Associates, 1976.

6. Horst, D. P., Tallmadge, G. K., and Wood, C. T. A Practical Guide to Measuring Project Impact on Student Achievement. Monograph No. 1 on Evaluation in Education. Washington, D. C.: U. S. Government Printing Office, 1975.

7. Becker, W. C. "Some Necessary Conditions for the Controlled Study of Achievement and Aptitude." In D. R. Green (ed), The Aptitude-Achievement Distinction. Monterey, California: CTB/McGraw-Hill, 1974.

8. Becker, W. C. "Some Effects of Direct Instruction Methods in Teaching Disadvantaged Children in Project Follow Through, Applications of Behavior Modification, Academic Press, Inc., 1975, pp. 139-159.

9. Metropolitan Achievement Tests Special Report, 1970 Edition. Report No. 1, June 1971, New York: Harcourt Brace Jovanovich, 1971, pp. 9-11.

Chapter 2

1. Loret, P. G., Seder, A., Biarchini, J. C., and Vale, C. A. Anchor Test Study: Equivalence and Norms Tables for Selected Reading Achievement Tests (grades 4, 5, 6), Office of Education Report No. 74-305. Washington, D. C.: U. S. Government Printing Office, 1974.

2. Becker, W. C. and Engelmann, S. Summary Analyses of Five-Year Data on Achievement and Teaching Progress with 14,000 Children in 20 Projects. Technical Report No. 73-2. Eugene, Oregon: University of Oregon Follow Through Project, December, 1973.

3. Ibid.

4. Education as Experimentation: A Planned Variation Model. Vol. III. Four-Year Effects of Follow Through. Cambridge, Mass.: Abt Associates, 1976.

5. Horst, D. P., Tallmadge, G. K., and Wood, C. T. A Practical Guide to Measuring Project Impact on Student Achievement. Monograph No. 1 on Evaluation in Education. Washington, D. C.: U. S. Government Printing Office, 1975.

Chapter 3

1. Education as Experimentation: A Planned Variation Model. Volume III. Four-Year Effects of Follow Through. Cambridge, Mass.: Abt Associates, 1976.

Chapter 8

1. Stallings, J. A. and Kaskowitz, D. H. Follow Through Classroom Observation Evaluation. Menlo Park, California: Stanford Research Institute, 1974.

2. McDaniels, G. L. The Evaluation of Follow Through. Educational Researcher, 1975, 4 (11), pp. 7-11.

3. Abt Associates, Education as Experimentation: A Planned Variation Model. Volume III. Four-Year Effects of Follow Through. Cambridge, Mass.: Abt Associates, 1976.

4. Kim, Y., Berger, B. J., and Kratochirl, D. W. Distar Instructional System. Palo Alto: American Institute for Research, 1972.

5. Science Research Associates. Summaries of Case Studies on the Effectiveness of the Distar Instructional System. Chicago: Author, 1971.

6. Becker, W. C., Engelmann, S. E. and Thomas, D. R. Teaching 1: Classroom Management. Palo Alto: Science Research Associates, 1975.