DOCUMENT RESUME

ED 116 135 CS 002 299

AUTHOR Kiesling, Herbert
TITLE The Relationship of Time Spent on Reading Instruction to Reading Gains as Measured by Norm-Referenced and Criterion-Referenced Tests.
PUB DATE 75
NOTE 32p.; Indiana University, Bloomington

EDRS PRICE MF-$0.76 HC-$1.95 Plus Postage
DESCRIPTORS *Achievement Tests; *Criterion Referenced Tests; Elementary Education; Grouping (Instructional Purposes); Individualized Reading; Norm Referenced Tests; Program Length; *Reading Instruction; Reading Research; *Reading Tests; *Time Factors (Learning)

ABSTRACT

This paper reports findings of research done concerning the productivity of types of classroom reading instruction in a sample of New York State elementary classrooms. Data was gathered from approximately 5,800 fourth, fifth, and sixth grade students in five New York State school districts during the spring semester of the 1974-75 school year. The main design of the analysis is based upon the "production function" paradigm used by economists, in which attention is paid to how the inputs of "ingredients" to a production process are mixed together to produce outcomes. To pinpoint instructional resources going into reading, teachers, aides, and specialists were asked to supply figures for minutes per week of reading instruction in each of four instructional modes: whole group, small group, individualized instruction, and individual help. Outcomes were measured by a standardized achievement test and a set of criterion-referenced tests. An important finding was that minutes of classroom teacher instruction seem to be related to student performances when measured by the criterion-referenced tests but not when measured by the standardized test. (MKM)

The Relationship of Time Spent on Reading Instruction to Reading Gains as Measured by Norm-Referenced and Criterion-Referenced Tests

Herbert Kiesling*

Introduction

This paper reports findings with respect to the productivity of types of classroom reading instruction in a large sample of New York State elementary school classrooms. The main design of the analysis is based upon the "production function" paradigm used by economists, in which attention is paid to how the inputs or "ingredients" to a production process are mixed together to produce outcomes. The resulting approach for the analysis in this study was similar to that suggested in recent years by some educational writers concerning the desirability of looking into the returns to instructional "time". Besides using a traditional standardized achievement test to measure outcomes, the study also utilized a set of criterion-referenced tests that have been developed by the New York State Education Department. This has made it possible to gain some important insights concerning the sensitivity of the two instruments. The most important finding was that minutes of classroom teacher instruction seem to be related to student performance when measured by the criterion-referenced tests but not when measured by the more traditional standardized test.

---

*Economics Department, Indiana University. The work reported upon here was part of the results of a joint effort by the author and Robert O'Reilly and Stephen Kidder of the New York State Education Department. Their assistance and cooperation is greatly acknowledged.

Part of the work discussed in this paper was financed by the National Commission on Productivity and part by the Carnegie Corporation of New York. Some of the work was done while the author was a member of the senior research staff at the Urban Institute.

The Paradigmatic Framework

The paradigm that economists use for viewing productivity questions is the "production function", which involves description of the possible relationships of the ingredients to a production process to the level of output (or outputs). It is virtually always true that the ingredients can be mixed in different ways to obtain the same amount of product, and since the ingredients usually have different costs, some mixtures that produce a given product level will be less expensive than others, i.e., they will be more efficient.*

This reasoning can be applied to a schooling process as well as to any other, and for reading instruction, for example, we could easily identify a number of important ingredients: three or four main types of teacher whole group instruction, specialist instruction, assistance by paraprofessionals, etc. Since all these will probably have different costs (some types of instruction consume more time, remuneration varies by type of instructor, etc.), different combinations of ingredients will imply different costs. Thus, if we are to satisfactorily deal with productivity issues in reading instruction, we need to obtain information on the relationship of the various instructional process inputs to student reading performance. All this assumes of course that a suitable measure of reading performance is available.

Since it is resource use that is our central concern within this paradigmatic framework, it is important that the instructional inputs to

---

*This is true for all but the most trivial of production processes where a single input is used to produce a single output.

For further discussion of the production function paradigm, see Bowles (1970), or Kiesling (1971a). Also any standard microeconomics textbook can be consulted.

be examined be quantified in a manner that is meaningful in resource-use terms. The way to do this that immediately suggests itself is to determine the time spent on each activity (in minutes per week, say) since this can be immediately converted to resource costs. (This assumes of course that we possess a fairly accurate notion of what is happening within elapsed time periods.*)

Study Objectives

The analytical approach just discussed was used to pursue two main research objectives. The first was to gain insights into the efficiency of the various alternative "factors of production" which go into the teaching of reading. Since this may vary according to type of student, part of this objective was to see whether apparent productivity patterns were similar for students from different socio-economic backgrounds.

The second objective was to compare the apparent effectiveness of the instructional inputs when reading progress of the same students was measured by the criterion-referenced tests and the norm-referenced test. As most educational researchers are fully aware, the latter type of test for basic cognitive subjects has provided the outcome measures for virtually all of the "educational production function" work of the past. Much of this work has shown both poor and inconsistent relationships between school inputs and student performance,** and many have wondered whether this may have

---

*In the past few years several educational writers who are not economists have also pointed out the value of quantifying school inputs in terms of time. See Bloom (1974), or Wiley and Harnischfeger (1974).

At least two empirical studies have been done which have quantified inputs in this way. One is a study by Conant (1973) who carefully traces the teaching productivity of teachers and paraprofessionals. The other is a study of compensatory reading instruction by the author (Kiesling, 1971b).

**See Averch, et al. (1974), pp. 46ff.

been due to imprecise measurement (vis-a-vis the objectives of given school districts) that may obtain when such tests are used.

The Measurement Instruments

The norm-referenced measurement instrument used in the study was the California Achievement Test, vocabulary and comprehension sections. The CAT is a widely respected commercially available test. Secondly, a set of criterion-referenced tests (CRT's) were used which have been developed in the New York State Education Department over the past several years.* The key difference between these tests and "standardized" tests (such as the California) is that test objectives are based upon the objectives of a much more circumscribed set of schools. There were two main reading aspects covered: vocabulary and comprehension. For the former, words were selected from the basal readers used by participating school districts. Comprehension objectives chosen were based in large part upon statements of teachers in the sample schools concerning their own priorities. Textual passages for comprehension exercises on the tests were also based upon passages in basal readers.

These tests also differed from norm-referenced tests in that tests for a given level contains material with a degree of difficulty that is applicable only to that level. (Levels were made to correspond to standard grades, using basal readers for guidance.) Mastery of a level implies getting nearly all items correct at that level. These tests are much more homogeneous than norm-referenced tests, which have questions at widely varying difficulty levels, many of which have no conceivable suitability

*The tests have been developed by the Bureau of School and Cultural Research, under the direction of Dr. Robert O'Reilly.

for measuring instruction in a given school grade.

Another important difference between the two types of test is that teachers are normally much more aware of the content of the criterion-referenced tests, not only because they participate in specifying objectives, but also because they are meant to have access to the tests before instruction takes place, since the tests themselves are meant to be an instructional aid. Teachers therefore are being constantly given timely feedback from prior testing. (The tests are administered often; every two to three weeks and teachers are given the results within five to ten school days.) Thus teaching to the tests is possible and teachers must be cautioned against teaching to specific items if long-run instruction is to be effective.*

Finally, it is only fair to state that these tests did not have one attribute that would have been desirable, which would be to measure only material taught in school and therefore where correct answers cannot be given based only upon intelligence. While it may or may not be possible to construct such a reading test, good scores on these criterion-referenced tests cannot be divorced from the advantages of native intelligence, although it is reasonable to expect that the intelligence component of the performance on these tests is less than would be the case for norm-referenced tests.**

*By all indications there has been little trouble with this in this New York state program, which is probably because participating teachers have been confident that they will never be rated in any way based on test results.

**For further discussion of the tests, see Athey (1975) and O'Reilly (1973).

The Data Set

The data were gathered for approximately 5800 4th, 5th, and 6th grade students in five New York state school districts during the spring semester of the 1974-75 school year. Four of the five school districts were urban core-cities, while the fifth district was suburban. The choice of districts was predicated largely on the fact that teachers in these districts were in substantial agreement concerning their teaching objectives, although the districts (and schools within them) varied in socio-economic characteristics. Students were administered the California Achievement Test in February and in May, and were also given four or five test administrations of the proper level of the CRT during the course of the semester, with test level being determined by the highest level on which the student missed a substantial number of questions.*

Other than test scores, the data were gathered using taped interviews with principals, teachers, specialists, and selected teacher aides. To meaningfully pinpoint instructional resources going into reading, teachers and specialists were asked to supply minutes per week of reading instruction in each of four instructional modes: whole group, small group, individualized

*Each test level corresponded roughly to a standard grade and each level had five parallel test forms each of which having between 30 and 40 items.

To choose a student's proper beginning level, the teacher was asked to make an estimate and the test for that level was administered. If that level turned out to be incorrect, the proper level was then readministered and became the beginning score. If a student scored medium to high on one level, he may have changed levels during the semester, but for comparability we have not included results from such level changes. Not many observations fell in that category, largely because each level spans an academic year and the instruction spanned less than half of an academic year.

instruction, and individual help.* An important virtue of the data set put together in this way was that it was possible to obtain estimates for each individual child.** When possible, estimates of one instructor were cross-checked against those of another. Project personnel were mature employees of the Bureau of School and Cultural Research and of the Urban Institute who were given careful training concerning interviewing techniques. Teacher interviews typically lasted one hour.

A number of other questions were asked in the teacher questionnaire and some data were gathered from school records. Such information included student age, absentee rate, father education, and father occupation. While project personnel did their best to get accurate information on the latter two variables from teacher, principal, and school records, it was still only possible to obtain it for about sixty percent of the students. Also gathered was considerable information concerning use of audio-visual and other instructional materials.

Finally, project personnel were not able to maintain desirable quality controls for the data collection effort for two schools (120

---

*These terms were used as follows. During whole group instruction the entire class is involved at the same time, listening to instruction, recitation, questions, etc. When sub-groups of the class are doing different tasks at the same time, the instruction was termed "small group." Individualized instruction involved diagnosis of individual needs and instruction tailored for those needs, while individual help involved drill-skill exercises with individual students.

**We decided to undertake such a difficult exercise only after some hesitation, but it did not turn out to be quite as difficult as it at first appeared. The method used was as follows. We presented a teacher with a class list of his or her children and asked first "What instruction did all receive together." Next we asked if there were any children whose type or time of instruction was different. If so, we then asked the teacher to name these students and then to supply the applicable information. If other instructors were involved, we then proceeded to ask the other instructor about the type and amount of instruction given each child by name.

children) in one school district and these data were discarded.*

The Explanatory Model

The statistical tool used for the explanatory models was single-stage multiple regression analysis. The variables included in the models, as well as the definition of the variables themselves, was motivated by the research objectives above. In order to inquire into the productivity questions posed, eight resource-using variables were included. Seven of these were the most important instructional time variables (on both a priori grounds and in terms of having meaningful variation in the data). The eighth was a variable constructed to denote the amount of materials and equipment that were used in the instruction.**

Perhaps the most important variable which must be accounted for in any determination of student post-test scores is their pre-test score. This helps to control for ability differences as well as for differences in prior preparation. Also, using pre-test scores as an explanatory variable, instead of using outright gains (subtracting the pre-test from the post-test score) allows for such systematic biases as the "regression to the mean" phenomenon to be dealt with by the model.***

*These were the only two schools representing a larger urban school district, and therefore only four school districts are represented in the study.

**Teachers were asked whether they used such equipment and materials as the following either occasionally or regularly: filmstrips, slides, overhead projectors, moving picture projectors, self-creativity materials, transparencies, charts, magnetic tapes, etc. The variable counted the number of positive responses to 14 such items, with the "regularly" response being weighted as double the "occasionally" response.

***However, the author has found in previous empirical work that findings using both approaches are highly similar. See Kiesling (1971), Appendix A, for a complete discussion.

It would be reasonable to expect on a priori grounds that student age would be related to reading performance, and this is even more true in the sample used in this study since students could be placed in a higher or lower CRT level according to their accomplishments. It was also possible to gather information concerning absentee rates, and this variable is included for obvious reasons.

Almost all educational production function models that have been used in the literature include a variable to account for the effects of student socio-economic background. The rationale for doing this is quite straightforward: parental influences are important for motivation towards learning as well as for how much learning takes place in the home directly. Since we were able to collect information on father occupation, it would seem sensible therefore that this be included in the explanatory model. However, there are some qualifications that should be made concerning the use of this variable.

First, we must remember that we already have a variable for pre-test score, and this can be expected to account for some of these home background effects. Secondly, it is true that single-stage multiple regression models of the educational process will be mis-specified no matter whether a SES variable is included or excluded.* A third qualification comes because in this study it was impossible for us to obtain data for more than one-third

*This is because SES is usually positively correlated with both school quality on the one hand, and pupil ability, motivation, and home learning on the other. Including the variable gives the family some of the credit that the school should get, while excluding it gives school variables too much credit. The truth lies somewhere in between. A properly specified model will use a simultaneous equation approach which explains both aspects with separate explanatory variables, but the data requirements for doing this are considerably richer than was available in the present study.

of the students, and we are not convinced that the one-third not collected are distributed randomly. As it turns out, introduction of the father occupation variable does not affect the findings very much, and we therefore use equations without the variable in a number of places.

Finally, an attempt was also made to examine socio-economic differences by stratifying the children according to whether they attended low and non-low socio-economic status schools.

Findings

Table 1 contains the principal set of fitted multiple-regression equations for all students (with the exception of the 120 children noted above.)* For reasons already explained, regressions are given with the father occupation variable both included and excluded. While there were six CRT levels used, results are presented for only five of them, since there were very few observations for students in the highest level (level 7). Levels 4, 5, and 6 are "proper" for the children in the study sample, since they were in grades 4, 5, and 6. Thus those children in levels 2 and 3 can be considered as being students who are below average in reading attainment.

The most important finding to be derived from the information in Table 1 is that student performance as measured on the norm-referenced test is much more poorly related to the instructional inputs than is such performance when measured by the criterion-referenced tests. Only one input shows signs of a positive relationship to CAT scores, (classroom teacher small group instruction) and even that does not reach statistical

*Means and standard deviations appear in Table 6.

Table 4
Results from Multiple Regression Equations Explaining Student Reading Performance in Grades 4, 5, 6, with and without Father Occupation Controlled

Explanatory Variables	CRT Level 2		CRT Level 3		CRT Level 4	
1. Classroom Teacher: Whole Group Instruction	.0031 (.167)	.0009 (.558)	.0091* (.000)	.0059* (.005)	.0005 (.759)	.0004 (.812)
2. Classroom Teacher: Small Group Instruction	.0045* (.075)	.0028* (.018)	.0029* (.058)	.0023* (.059)	.0021* (.091)	.0021* (.040)
3. Classroom Teacher: Individualized Instruction	.0056 (.543)	.0100 (.147)	.0106 (.163)	.0088* (.062)	.0017 (.696)	.0022 (.386)
4. Specialist: Small Group Instruction	.0087 (.567)	-.0034 (.725)	.0066 (.502)	.0132 (.121)	.0361 (.115)	.0149 (.267)
5. Specialist: Individualized Instruction	.0017 (.532)	-.0021 (.326)	-.0019 (.535)	-.0027 (.293)	-.0079 (.023)	-.0088 (.003)
6. Paid Aide: Small Group Instruction	-.0012 (.928)	.0017 (.815)	-.0014 (.756)	-.0025 (.325)	-.340 (.001)	-.0241 (.001)
7. Individual Help	-.0017 (.773)	.0007 (.890)	-.0013 (.839)	-.0005 (.923)	-.0058 (.383)	-.0086 (.059)
8. Material and Equipment Index	.0495 (.762)	-.1144 (.311)	-.2113 (.059)	-.1926 (.021)	-.3429 (.000)	-.3597 (.000)
9. Student Age	.0320 (.815)	.0115 (.915)	-.1714 (.146)	-.1517 (.088)	.0185 (.838)	-.1216 (.069)
10. Number of Days Absent	-.0371* (.059)	-.0210 (.142)	.0031 (.838)	.0015 (.905)	.0194 (.168)	.0067 (.524)
11. Father Occupation	-.0071 (.984)		.3559* (.057)		.1357 (.355)	
12. Pre-test Score	.708* (.000)	.608* (.000)	.583* (.000)	.543* (.000)	.733* (.000)	.728* (.000)
Number of Observations	153	260	320	507	438	658
R^2	.542	.429	.363	.304	.599	.579

Table 1
(continued)

Explanatory Variables	CRT Level 5		CRT Level 6		CAT	
1. Classroom Teacher: Whole Group Instruction	.0013 (.576)	.0033 (.129)	.0036 (.109)	.0024 (.236)	-.0008 (.738)	-.0004 (.984)
2. Classroom Teacher: Small Group Instruction	-.0012 (.529)	-.0012 (.435)	.0040* (.050)	.0035* (.051)	.0019 (.267)	.0017 (.209)
3. Classroom Teacher: Individualized Instruction	.0010 (.834)	-.0009 (.806)	.0071* (.088)	.0057 (.128)	.0012 (.816)	-.0038 (.288)
4. Specialist; Small Group Instruction	.0141 (.492)	-.0121 (.412)	-.7497 (.014)	-.0065 (.025)	-.0446 (.011)	-.0355 (.002)
5. Specialist: Individualized Instruction	.0055 (.743)	-.0025 (.785)	.0074* (.012)	.0752* (.012)	-.120 (.017)	-.0123 (.001)
6. Paid Aide: Small Group Instruction	-.0041 (.790)	-.0088 (.142)	.0440 (.137)	.0515* (.097)	-.0213 (.015)	-.0119 (.025)
7. Individual Help	.0185 (.198)	.0125 (.344)	.0626* (.099)	.0460 (.191)	-.0194 (.016)	-.0126 (.030)
8. Material and Equipment Index	-.4066 (.006)	-.3061 (.008)	-.4206 (.000)	-.2771 (.010)	.1585 (.166)	.0560 (.520)
9. Student Age	-.1920 (.115)	-.1227 (.254)	.2087 (.145)	.1671 (.217)	.0555 (.630)	-.1085 (.232)
10. Number of Days Absent	-.0088 (.645)	-.0253* (.099)	.0137 (.353)	.0027 (.847)	-.0528* (.012)	-.0635* (.000)
11. Father Occupation	.5228* (.004)		.2538 (.103)		1.507* (.000)	
12. Pre-test Score	.713* (.000)	.716* (.000)	.713* (.000)	.712* (.000)	.863* (.000)	.905* (.000)
Number of Observations	436	582	397	472	2231	3172
R^2	.614	.558	.674	.623	.790	.797

Table notes appear after Table 5.

Table 2
Results for Type of Instruction, All Students

Minutes per week Instruction by:	CRT Level 2	CRT Level 3	CRT Level 4	CRT Level 5	CRT Level 6	CAT
Classroom Teacher	.0020	.0029*	.0018*	-.0003	.0038*	.0009
	(.165)	(.010)	(.066)	(.831)	(.020)	(.480)
Reading Specialist	-.0025	-.0006	-.0067	-.0094	.0120	-.0142
	(.242)	(.803)	(.017)	(.202)	(.299)	(.000)
Paid Paraprofessional Aides	-.0013	-.0017	-.0136	-.0046	.0507	-.0141
	(.775)	(.467)	(.000)	(.409)	(.102)	(.001)
Unpaid Teacher Aides	-.0011	-.0090	-.0370	b.	.0786	-.0118
	(.188)	(.482)	(.089)		(.129)	(.418)

Table Notes appear after Table 5.

Table 3
Effect on Main Instructional Variances of Including
Dummy Variables for School District

Explanatory Variable	CRT Level 2		CRT Level 3		CRT Level 4	
1. Classroom Teacher: Whole Group Instruction	.0009 (.538)	.0011 (.559)	.0059* (.005)	.0066* (.002)	.0004 (.812)	.0019 (.232)
2. Classroom Teacher: Small Group Instruction	.0028* (.018)	.0019 (.353)	.0023* (.059)	.0026* (.044)	.0021* (.040)	.0025* (.026)
3. Classroom Teacher: Individualized Instruction	.0100 (.147)	.0078 (.256)	.0088* (.062)	.0082* (.0807)	.0022 (.386)	.0008 (.775)
4. Specialist: Individualized Instruction	-.0021 (.326)	-.0034 (.125)	-.0027 (.293)	-.0045 (.082)	-.0088 (.003)	-.0103 (.000)
5. Dummy Variable: District A		0.191 (.812)		-0.238 (.633)		0.316 (.461)
6. Dummy Variable: District B		-0.732 (.428)		-2.681* (.000)		-1.694* (.000)
7. Dummy Variable: District C		-2.875* (.007)		-2.217* (.002)		-2.212* (.000)

Table Notes Appear After Table 5

Table 3
(continued)

Explanatory Variable	CRT Level 5		CRT Level 6		CAT	
1. Classroom Teacher: Whole Group Instruction	.0033 (.129)	.0025 (.238)	.0024 (.236)	.0014 (.454)	-.00004 (.984)	-.0031 (.092)
2. Classroom Teacher: Small Group Instruction	.0012 (.435)	-.0006) (.696)	.0035* (.051)	.0039* (.024)	.0017 (.209)	-.0003 (.836)
3. Classroom Teacher: Individualized Instruction	-.0009 (.806)	-.0083 (.021)	.0057 (.128)	.0008 (.837)	-.0038 (.288)	-.0007 (.851)
4. Specialist: Individualized Instruction	-.0121 (.412)	-.0151 (.090)	.0752* (.012)	.0429 (.121)	-.0123 (.001)	-.0085 (.018)
5. Dummy Variable: District A		-0.608 (.282)		-1.386* (.027)		0.563 (.312)
6. Dummy Variable: District B		-1.329* (.022)		-3.708* (.000)		6.217* (.000)
7. Dummy Variable: District C		-4.791* (.000)		-6.177* (.000)		4.451* (.000)

Table 4

Comparative Results for Students in Low and Non-Low SES Schools

Explanatory Variable	CRT Level 2		CRT Level 3		CRT Level 4	
	Low	Non-Low	Low	Non-Low	Low	Non-Low
1. Classroom Teacher: Whole Group Instruction	.0022 (.261)	a.	.0050* (.043)	.0052 (.292)	.0024 (.244)	-.0063 (.043)
2. Classroom Teacher: Small Group Instruction	.0019 (.351)		.0017 (.256)	.0005 (.875)	.0041* (.003)	-.0012 (.600)
3. Classroom Teacher: Individualized Instruction	.0378* (.003)		.0067 (.459)	.0024* (.006)	.0035 (.250)	.0039 (.488)
4. Specialist: Small Group Instruction	-.0095 (.370)		.0033 (.759)	.0064 (.644)	.0088 (.596)	.0284 (.232)
5. Specialist: Individualized Instruction	-.0044 (.183)		-.0008 (.838)	-.0260 (.004)	-.0129 (.014)	-.0246 (.245)
6. Paid Aide: Small Group Instruction	.0121 (.351)		.0011 (.667)	-.0157 (.205)	-.0231 (.001)	.0956 (.383)
7. Individual Help	-.0016 (.748)		.0044 (.540)	-.0079 (.257)	-.0027 (.603)	.0035 (.848)
8. Materials and Equipment Index	-.822 (.010)		.0097 (.952)	.2123 (.395)	-.231 (.022)	-.0459 (.769)
9. Student Age	-.0233 (.857)		-.124 (.286)	-.1067 (.506)	-.233 (.005)	-.1123 (.369)
10. Number of Days Absent	-.0233 (.243)		-.0027 (.890)	-.0209 (.550)	-.0022 (.862)	.0636 (.011)
11. Schools: Mixed Black and White (dummy variable)	-0.641 (.515)		1.017* (.094)	0.913 (.476)	1.533* (.002)	-2.807 (.201)
12. Schools: Mostly Black (dummy variable)	3.288* (.018)		-1.157 (.177)		-.395 (.455)	
13. Pre-Test Score	.471* (.000)		.395* (.000)	.536* (.000)	.627* (.000)	.754* (.000)
Number of Observations	174		275	179	386	239
R^2	.387		.230	.338	.547	.526

Table Notes appear after Table 5.

Table 4.
(continued)

Explanatory Variable	CRT Level 5 Low	CRT Level 5 Non-Low	CRT Level 6 Low	CRT Level 6 Non-Low	CAT Low	CAT Non-Low
1. Classroom Teacher: Whole Group Instruction	.0022 (.414)	.0018 (.637)	.0041 (.300)	-.0017 (.682)	.0014 (.510)	.0021 (.587)
2. Classroom Teacher: Small Group Instruction	.0011 (.607)	-.0041 (.152)	.0068* (.019)	.0066* (.038)	.0043* (.019)	-.0011 (.672)
3. Classroom Teacher: Individualized Instruction	.00005 (.994)	-.0109 (.072)	.0034 (.589)	.0054 (.540)	.0063 (.198)	-.0201 (.000)
4. Specialist: Small Group Instruction	-.0139 (.380)	-.0021 (.951)	-.383 (.335)	b.	-.0221 (.063)	-.0413 (.081)
5. Specialist: Individualized Instruction	-.0070 (.446)	b.	.0474 (.252)	.1097* (.008)	-.0051 (.245)	-.0127 (.325)
6. Paid Aide: Small Group Instruction	-.0069 (.264)	-.0171 (.405)	b.	.0448 (.121)	-.0100 (.048)	.0068 (.767)
7. Individual Help	.0009 (.958)	.0173 (.389)	-.0585 (.365)	.1032* (.031)	-.0108 (.130)	-.0221 (.016)
8. Materials and Equipment Index	-.349 (.040)	-.489 (.035)	.1602 (.380)	-.2465 (.263)	-.2702 (.100)	.1826 (.339)
9. Student Age	-.135 (.380)	-.1954 (.223)	.554* (.023)	.1197 (.521)	-.116 (.320)	.0379 (.798)
10. Number of Days Absent	-.0321 (.143)	-.0183 (.558)	-.0278 (.211)	-.0325 (.270)	-.0426* (.035)	.0064 (.849)
11. Schools: Mixed Black and White (dummy variable)	1.513* (.016)	1.288 (.233)	2.637* (.003)	0.282 (.875)	-4.312* (.000)	5.155* (.000)
12. Schools: Mostly Black (dummy variable	2.494* (.012)		2.080* (.023)		-1.451 (.070)	
13. Pre-Test Score	.670* (.000)	.659* (.000)	.584* (.000)	.761* (.000)	.877* (.000)	.909* (.000)
Number of Observations	288	276	190	248	1458	1443
R^2	.582	.536	.622	.708	.801	.780

Table 5
Results for Type of Instructor:
Students in Low and Non-Low SES Schools

Minutes per week Instruction by:	CRT Level 2		CRT Level 3		CRT Level 4	
	Low	Non-Low	Low	Non-Low	Low	Non-Low
Classroom Teacher	.0027*	a.	.0023*	.0024	.0024*	-.0003
	(.096)		(.093)	(.355)	(.035)	(.894)
Reading Specialist	-.0058		-.00004	-.0109	-.0053	.0032
	(.066)		(.992)	(.078)	(.199)	(.844)
Paid Paraprofessional Aide	.0032		.0011	-.0095	-.0111	.0052
	(.548)		(.665)	(.074)	(.006)	(.813)
Unpaid Teacher Aides	-.0131		-.0012	-.0072	-.0057	-.0646
	(.154)		(.538)	(.657)	(.866)	(.016)

TABLE NOTES

The figure in parentheses under the coefficent of net regression denotes the level of probability with which it could occur that the result occured by chance. A figure of .050 means the probability is 5 chances in 100 for example. Coefficients significant at the ten-percent level or lower and which have the hypothetically correct sign are marked with an asterisk.

a. There were too few observations in the population for this regression equation to be meaningful.

b. No observation had a non-zero value for this variable.

Table 5
(continued)

Minutes per week Instruction by:	CRT Level 5		CRT Level 6		CAT	
	Low	Non-Low	Low	Non-Low	Low	Non-Low
Classroom Teacher	.0026	-.0044	.0050*	.0054*	.0040*	.0011
	(:168)	(.068)	(.017)	(.068)	(.009)	(.692)
Reading Specialist	-.0047	-.0032	.0015	.1132*	-.0107	-.0260
	(.522)	(.928)	(.229)	(.007)	(.008)	(.019)
Paid Paraprofessional Aide	-.0051	-.0046	b.	.0430	-.0106	-.0194
	(.393)	(.746)		(.142)	(.013)	(.066)
Unpaid Teacher Aides	b.	b.	b.	.0874*	-.0172	-.0077
				(.083)	(.345)	(.731)

Table 6
Description of Variables

Variable	$\bar{m}$	$\bar{s}$
Whole Group Instruction, Classroom Teacher, Minutes Per Week (MPW)	64.07	102.59
Small Group Instruction, Classroom Teacher, MPW	166.35	140.79
Small Group Instruction, Specialists, MPW	2.16	15.79
Small Group Instruction, Paid Aides, MPW	4.89	33.85
Individualized Instruction, Classroom Teacher, MPW	11.50	51.07
Individualized Instruction, Specialists, MPW	11.91	50.61
Whole Group Instruction, Total MPW	64.07	102.59
Small Group Instruction, Total MPW	173.65	146.88
Individualized Instruction, Total MPW	24.22	72.82
Individual Help, Total MPW	7.85	30.16
Classroom Teacher Instruction, Total MPW	244.31	146.18
Specialist Instruction, Total MPW	14.32	54.07
Paid Aide Instruction, Total MPW	9.75	43.60
Unpaid Aide Instruction, Total MPW	1.42	12.22
Criterion Referenced Test, (CRT) Level 2, Test 1 (February)	12.72	5.05
CRT Level 2, Test 4 (May)	13.22	5.05
CRT Level 3, Test 1	13.53	4.12
CRT Level 3, Test 4	14.57	4.45
CRT Level 4, Test 1	16.09	4.81
CRT Level 4, Test 4	16.68	5.23

Table 6 (continued)

Variable	$\bar{m}$	$\bar{s}$
CRT Level 5, Test 1	18.07	5.99
CRT Level 5, Test 4	18.60	6.24
CRT Level 6, Test 1	18.59	6.01
CRT Level 6, Test 4	19.01	5.95
California Achievement Test, Pre-Test (February, in Months)	50.18	21.39
California Achievement Test, Post-Test (May, in Months)	55.06	22.09
Index of Materials Used	3.49	2.10
Student Age (Half Years)	21.83	2.04
Number of Days Absent, Student	10.75	11.51
Father Occupation Index	3.48	1.24

significance at the twenty percent level. Indeed, most of the instructional inputs are negatively related to student performance on the CAT and this often occurs at advanced levels of statistical significance. This is expecially true for non-classroom teacher instruction.

The positive relationship of school instructional inputs to the CRT scores is especially noteworthy for instruction by the classroom teacher, particularly when it is in the small group mode. However, specialist instruction shows signs of being important in levels 4 and 6, while in level 5 no input seems to be related to student performance.

There is an important reservation that must be made concerning the relationships to these criterion-referenced test findings however, which is not directly shown in Table 1. Although many instructional inputs were significantly related to student performance measured in the CRT's, it is also true that the amount of variation explained was not large. Thus, more than 300 minutes of classroom teacher time would be required for a one point gain in CRT score. This is about three standard deviations of the teacher whole class instruction variable. For small group instruction more than 400 minutes would have been required, again about three standard deviations. (There were typically 30 items per CRT level, corresponding to an academic year.) By contrast, apparent pupil average gain as measured by the standardized measure was more than .4 grade equivalents. Can the difference be explained as due to the difference of the quantity of home learning as opposed to school induced learning?

Despite our theoretical expectations about the influence of a variable for parents' occupation, its inclusion in the explanatory equation does not seem to alter the findings very much, and since more than one-third of the observations have this information missing, the variable is not included

in the fitted equations presented in the rest of the paper. However, it is quite noteworthy that father's occupation displays a much stronger relationship to performance on the CAT than on the CRT's, which suggests that the latter are more sensitive to learning originating in the school as opposed to that in the home, just as we would expect a priori. This would even be more true of course for learning outside the proper class level, which is captured by standardized tests and (correctly, from the standpoint of studying schools) not captured by the CRT's.

One finding contained in Table 1 is both puzzling and potentially important, and that is the consistently negative relationship of the amount of materials and equipment used to student performance as measured by the CRT's. A partial explanation might be the fact that the lower SES schools have more equipment, perhaps purchased with Title I money.* Otherwise, the use of equipment and materials would seem to be counterproductive for some mysterious reason. The story becomes even more perplexing when we notice that for student progress as measured by the California Achievement Test the materials and equipment variable reverses sign and, when father occupation is controlled, even approaches statistical significance.

Another result shown in the table which is of interest has to do with the marked differences between the relationships of the instructional inputs to student performance in CRT levels 5 and 6. One possible explanation for this is that level 5 has more than its share of older students with low ability. Evidence to support this is contained in the coefficients for student age in the two levels, positive in level 6 and negative in level 5.

---

*The index value for low SES schools was 3.92 and for non-low schools 3.59.

Finally, some explanation is required for the number of statistically significant negative relationships for the specialist, paid aide, and individual help instructional inputs. A very likely explanation for this is that such instruction is generally devoted to "remediation" for low achieving students. This is especially likely in this data set because of the low amounts of instructional time that were devoted to these modes.*

It appears in Table 1 that regular instruction by the classroom teacher is more important for student performance than for instruction by other types of instructional personnel. This is confirmed in Table 2, where results are given for the same model as in Table 1 (without the father occupation variable) except that minutes of all type instruction by type of instructor is used. Except for level 5, the only significant or near-significant net regression coefficients are for the classroom teacher. These findings may be due in part to the strong possibility mentioned above that other types of instructors were used for lower ability students.

It was also possible to test whether there were meaningful differences between type of instruction: whole-group versus small-group, etc. There were none, and these findings are omitted.

In our sample there were only four usable school districts, and with such a small number it is perhaps not very meaningful to inquire into school district effects. However, it is possible to at least ascertain whether school district per se seems to be an important variable by merely including dummy variables for each district. The results when this is done, for the main instructional variables, are presented in Table 3. The dummy variables show significant differences by school district more often

*See the means presented in Table 6.

than not, but, except for at CRT level 2, including the dummy variables does not affect the explanatory power of the instructional variables.

Perhaps the most intriguing result in Table 3 is the rather spectacular reversal in sign and significance for the dummy variables for districts B and C when reading performance is measured by the CAT. No ready explanation for this comes to mind, but again it happens that results for the two types of test are highly divergent.

The last hypothesis tested in the study had to do with whether instructional results differ for high and low SES pupils and whether results seemed to be much different when there were large percentages of black children in a given school. To do this each school was placed in a "low-SES" or "non-low SES" category, and schools were also classified as to whether all-white, all-black, or racially mixed. Then the regression explanatory model was fitted to the performance of children in the low and non-low SES groupings, and within each of these categories dummy variables were introduced to account for racial mixture. These findings are presented in Tables 4 and 5. The hypothesis that instructional relationships differ for the low and non-low SES schools is confirmed by the information in these tables, with instruction being more closely related to the gains of low SES pupils. This even holds true with respect to performance as measured by the California Achievement Test, which is remarkable in that it is the only instance in the entire study where this is true. This close relationship for the low SES students may reflect the fact that due to the nature of their homes a much greater percentage of their performance gains are due to school instruction than is true for the high SES children.

Finally, while the results of race in the findings in Table 4 seem somewhat mixed, in general the all-black schools normally show up as

performing better on the CRT tests and worse on the NRT. The implication of this relationship is that the poorer showing for blacks in many studies using norm-referenced tests is indeed due to their poorer home environments, and when this is properly adjusted for (or when an instrument that measures school effects is used) blacks do as well or better than whites. However these findings are based on so few school observations that it would be wrong to generalize them very far.

Cost Ramifications

It is perilous to present cost ramifications (and therefore direct efficiency estimates) of findings from a single study where the effective instruction lasted only eight to ten weeks.* However, the delineation of the procedure for arriving at cost estimates should not be without interest, nor should a set of illustrative calculations be without useful information.

It is not very difficult to derive meaningful cost estimates. First the calculated coefficient of net regression is divided into 1.0. This shows how many minutes of instruction it will take to raise pupil scores one point. This is then divided by 1380, the figure taken is the total number of instructional minutes in a teacher week.** This figure is then multiplied by the instructor's yearly salary and divided by class size to

*What would be needed before policy decisions could be directly based upon such calculations would be (1) carefully controlled variation of inputs, (2) good control for student socio-economic characteristics (3) studies spanning a time period of at least a year and preferably longer, and (4) a number of replications.

**This figure comes from a finding by Educational Turnkey Systems (1971) that teachers spend 4.6 hours per day doing instruction.

yield the cost per student per point gain.* The only inputs that were consistently significant enough across most of the CRT levels to merit making calculations were whole-group and small-group instruction by the classroom teacher. Taking average coefficients across all levels (and using equations both with and without father occupation), the figures obtained were $87 and $431 cost per point gain per student for whole-group and small-group instruction respectively. For the CAT results, only two coefficients were positive, those for small group instruction and the materials index. Calculations for the cost for .1 grade equivalents of gain for these two inputs came to $575 and $502 respectively.**

Finally, the only student population where a fair number of inputs were positively related to reading performance was that in CRT level 6. Therefore the cost calculations for level 6 inputs were computed and are as follows:

Teacher whole-group instruction	$ 89
Teacher small-group instruction	$ 276
Teacher individualized instruction	$1132
Specialist small group instruction	$ 117
Paid aide small group instruction	$ 9

*The median figures in our data for class size and small group size were 27.0 and 7.0 respectively. These were the figures used. $10,000 and $12,000 were the figures used for the annual salaries of classroom teachers and specialists respectively. Paid aides were assumed to receive $5.00 per hour, including fringes.

**Calculations for costing out the materials index are not difficult, merely somewhat tedious. Needed are data concerning the cost of each item, its normal life, and how many minutes per week it is used. This information can be used to construct a cost-weighted materials index if desired. The index used in the regressions was not weighted and therefore an average value for each item needed to be taken.. Assumptions were also made that regular use meant half time and occasional use one-fourth time, a procedure which undoubtedly overstated the cost. However, results for the materials index were so indifferent that it was not considered worthwhile to spend the considerable time it would have taken to make its costing precise.

Obviously the use of para-professionals, if they can be made to be instructionally effective, is quite cost-effective. Otherwise, teacher whole-group instruction has the least cost.

Concluding Comments.

The data analysis in this study has yielded important findings with respect to three lines of inquiry. First, it was asked whether some school inputs are more effective in teaching reading than others. It was found that the activities of the regular classroom teacher seem to be consistently more important, although other inputs were sometimes related to performance, especially in CRT level 6. Next was asked whether the relationships of school inputs to reading performance as measured by criterion-referenced tests might not be different from those measured by norm-referenced tests. It was found that the former measure was related to school inputs much more consistently. Finally, we asked whether input-performance relationships might not be different for students from different socio-economic backgrounds. When students were separated according to whether they attended low or non-low socio-economic status schools, it was found that school inputs were more consistently related to low SES student performance.

While it would be dangerous to conclude too much from the findings of a single study which monitored only about nine or ten weeks of actual instruction, these findings are still quite suggestive. After all the non-findings that have been turning up in the empirical multi-variate literature concerning the effect of the classroom teacher, it is reassuring to see some more positive results using tests which are perhaps more sensitive than those used in that literature. However, even though the instruction period was quite short, the average student gains on the criterion-referenced tests are surprisingly low, only about one point on a

thirty-point instrument (built to cover one school year). On the other hand, students gained more than four-tenths of a year measured in standardized grade equivalents. Could we have succeeded in distinguishing between home effects and school effects? Or at least we may have distinguished between instruction proper to the grade level and instruction that may have somehow given the students skills which are "proper" for other grade levels.

Finally, we must make an important reservation about the lack of positive relationships to reading performance of instruction by specialists and para-professionals. Two things can be said about this. First, it is a fact that in this sample of classrooms non-classroom teacher instruction was simply not used very much.* It can be argued that instruction will only begin to take hold after it reaches some threshhold, and that for almost all the cases in this sample it is difficult to see how such a threshhold could have been reached. Secondly, one strongly suspects that instruction by non-classroom teachers is typically used for children who are "troublesome"; who have traits both in lack of intelligence and motivation which could not have been controlled for by the relatively crude non-instruction variables used in this study. Therefore no conclusions about the effectiveness of the inputs of such personnel should be made until such time as their effectiveness can be examined using more controlled experiments.**

---

*See Table 6. Specialist small group instruction averaged only two minutes per week, paid-aide small group instruction five minutes per week, and specialist individual instruction 12 minutes per week. Of the 5887 students in the total sample, 5717 had no small group specialist instruction, 5702 no small group paid aide instruction, 5340 had no specialist individualized instruction, and 5830 had no paid aide individualized instruction.

**This statement is reinforced by the fact that the author found strong, positive relationships between specialist instruction and reading gains by Title I children in an earlier study of California Title I projects. See Kiesling (1971b, 1972).

References

Athey, Irene, "Models for Measuring Growth and the Measurement of Outcomes," Paper presented at the Meetings of the American Educational Research Association, Washington, D.C., 1975.

Averch, H., et al., How Effective is Schooling?, Educational Technology Press, Englewood Cliffs, N.J., 1974.

Bowles, S., "Towards an Educational Production Function," In L. Hansen, ed., Education, Income, and Human Capital, Columbia University Press, New York, 1970.

Bloom, B.S., "Time and Learning," American Psychologist, September 1974, pp. 682-688.

Bureau of School and Cultural Research, New York State Education Department, Measurement and Productivity in School Reading Programs: Main Report, Albany, N.Y., 1975.

Conant, Eaton H., Teacher and Paraprofessional Work Productivity, Lexington Books, Lexington, Mass., 1973.

Educational Turnkey Systems, Inc., "Economics of Third Grade Education," Audio-Visual Instruction, November 1971.

Kidder, S., O'Reilly, R., and Kiesling, H., "Quantity and Quality of Instruction: Empirical Investigations," Paper presented at the American Educational Research Association Meetings, Washington, D.C., 1975.

Kiesling, H.J., "Multivariate Analysis of Schools and Educational Policy," P-4595, The Rand Corporation, Santa Monica, California, 1971a.

Kiesling, H.J., "Input and Output in California Compensatory Education Projects," R-781-CC/RC, The Rand Corporation, Santa Monica, California, 1971b.

Kiesling, H.J., "Reading Performance of Disadvantaged Children: Cost Effectiveness of Educational Inputs," Education and Urban Society, November, 1972.

Kiesling, H., "Broad Results for Instructional Variables and Cost Estimates," Paper presented at the American Educational Research Association Meetings, Washington, D.C. 1975.

O'Reilly, R., "The Development of the New York State Bank of Reading Objectives," (mimeo), The New York State Education Department, Albany, N.Y., 1973.

O'Reilly, R., "The Contributions of Quantity and Quality of Instruction to Reading Programs," Paper presented at the American Educational Research Association Meetings, Washington, D.C., 1975.

Wiley, D. and Harnischfeger, A., "Explosion of a Myth: Quantity of Schooling and Exposure to Instruction, Major Educational Vehicles," Educational Researcher, April 1974, pp. 7-12.

Winkler, D.R., "Time and Learning: An Economic Analysis," (mimeo), Economics Department, University of California at Santa Barbara, 1975.