Harvard Educational Review

VOLUME 48 NUMBER 1 1978

A New Perspective on the Effects of First-Grade Teachers on Children's Subsequent Adult Status

EIGIL PEDERSEN
McGill University
THÉRÈSE ANNETTE FAUCHER
Ministère de l'Education du Québec
with WILLIAM W. EATON
McGill University

In this article Eigil Pedersen, Thérèse Annette Faucher, and William W. Eaton have taken on one of the most difficult questions in educational research: the impact of the classroom teacher on children's adult status. The authors detail the results of a research project of unconventional methodology and unusually long duration. They sought originally to examine atypical IQ changes but came to focus on the enduring effects of one remarkable first-grade teacher. Drawing upon the idea of the "self-fulfilling prophecy," the authors relate the effects of teachers' attitudes and resultant behavior to the subsequent adult status of sixty children. Their conclusion, in contrast to that of many recent studies, is that the classroom teacher may have a significant effect on children's chances for success in later life.

Social-science and educational-research textbooks frequently describe and recommend logical and orderly procedures for conducting research, and the designers of many research projects probably follow these directions. Despite these recommendations, real-life circumstances frequently impose conditions that make routine methods of research impossible. We are convinced, nonetheless, that useful work can be done by employing the ad hoc procedures often demanded by less-than-ideal circumstances. In the pages that follow, we report the findings of one non-standard project in the hope that these data will illustrate a potentially effective approach to educational research and, at the same time, illuminate one of the

Harvard Educational Review, Vol. 48, No. 1, February 1978.

0017-8055/78/0200-0001$02.42/0

most obscure problems in the field: the impact of classroom teachers upon children's subsequent adult status.

By every standard our research has proceeded in a most atypical fashion. First, one of the authors[1] has been personally involved with the research setting since the age of four, initially as a pupil and later as a teacher. Hence, the choice of research questions resulted from reflections upon personal experiences rather than from a standard review of the scholarly literature. Another unusual aspect is that the research was conducted in short episodes over a long period of time—not only in the early exploratory stages but also in the stages of analysis and writing. Progress was slow because the author's heavy administrative responsibilities made him very much the part-time researcher. There was ample opportunity, therefore, to become aware of the changing concerns and the state of knowledge in this field during the long period between 1967, when the first parts of the data came to the attention of the author, and the present time. This undoubtedly led to somewhat different findings than might have emerged from a more conventional research project. As a result, the direction of this study was greatly influenced by three unexpected, if not serendipitous, findings: the discovery, while checking permanent record cards for pupils' addresses, of a regular pattern of IQ change; the discovery that IQ change varied by first-grade teacher and by the gender of the student, a finding that led us to focus on the effect of teachers rather than IQ change; and the discovery of an unexpected relationship between first-grade teachers and adult status. However tortuous the path has been, the results have been dramatic, as we will attempt to demonstrate.

Our research setting, the Ray School (a fictitious name), was located in a large northeastern city in North America. Situated in one of the poorest areas of the city, the fifty-year-old building that housed its students stood out like a fortress in the streets. During the period when the subjects attended Ray, freight terminals of a large railway, as well as a steel-fabrication plant, were located in its immediate neighborhood, and most of the pupils had to cross at least one major traffic artery to get to the school. Across the street from the front entrance, the buildings of a brothel, thinly disguised as residences, blocked the view of a junkyard. Crowded tenement houses were interspersed with an automobile repair shop, a dry-cleaning plant, and an armature-wiring factory. The asphalt schoolyard was enclosed by a chain-link fence and the ground-floor windows were protected with vertical iron bars.

At the time the subjects were attending Ray, approximately one-third of the students were Black children; the rest were mainly Anglo-Saxon in origin, with a sprinkling of pupils from other ethnic backgrounds. All of the 400 to 500 pupils were at least nominally Protestant. Given the mores of the time, boys and girls had their own entrances, used separate basements, and were segregated into different play areas in the yard. Although classes were integrated by gender, children moved from place to place in formal line at the sound of bells, girls in front and boys behind. The children were frequently unruly and often fought with one another. Certain children, moreover, were often strapped—officially there

1 When the word "author" is used, it refers to the senior author.

were about 500 strappings a year—and occasionally a difficult child was referred to a detention home.

Ray had the reputation among teachers as being the most difficult school under its particular board. Not surprisingly, teachers reluctantly accepted assignments there, and a rapid and continuous turnover of novice principals and newly trained teachers was commonplace. Nonetheless, a solid core of experienced teachers continued to teach at the Ray School since the Depression.

In the total system of about eighty elementary schools of which Ray was a part, it was the school board's policy to administer group IQ tests routinely to all third- and sixth-grade children. Compared to others in the system, Ray School students consistently had the lowest mean IQ.

When a class was promoted into the following grade, it was generally the policy to divide it into two equivalent groups so that each of the two teachers at the next grade received a fair share of boys and girls, of promising and slow students, and of well-behaved children and "troublemakers." Additionally, the school had two "special" classes—one junior and one senior—to accommodate the supposedly small proportion of children who could not benefit from placement in heterogeneous classes. This process of grading and dividing classes was continued throughout each child's entire elementary-school career.

When the author returned as a teacher to the Ray School, in the early 1950s, it was using a two-track system, with two equal classes in each grade. Because of shared teaching with the other fifth-grade teacher and because of many extracurricular activities, he became well acquainted with the seventy-five children in both classes. He followed the progress of these children over several grades, maintaining an interest in their subsequent academic achievement. He was shocked by two things: the widespread failure by Ray School graduates, even scholarship winners, in the first year of high school and the extremely poor high-school completion rate. Only five of the seventy-five students ultimately graduated from high school. Even in light of the low IQ scores previously mentioned, this poor record of achievement ultimately motivated the author to embark on this research study, convinced that different outcomes in school achievement must be related to factors over and above differences in IQ.

Because all social research should have well-defined boundaries, it was necessary to decide which factors to study. For years the author had been critical of the amount of time and effort invested in the discovery of unalterably gloomy facts associated with the high drop-out rate of disadvantaged pupils. While recognizing the importance of the pursuit of knowledge for its own sake, he decided to concentrate on some aspects of the problem that might be subject to change as a result of intervention. The research, in other words, was intended not simply to explain failure, but to produce findings that might lead to improvements of the academic performance of future students.

The Research Project

A major part of the data for this project came from the permanent record cards of students attending the Ray School. Each card contained a cumulative record of

family background, school achievement, and anecdotal information about the children during their seven years at the school. Each card noted the child's date of entry, the class to which he or she was admitted, and considerable personal information, including birth date, records of vaccination, address, parents' names, parents' occupations, telephone numbers, names of brothers and sisters, and additional information considered potentially useful to the school. During the child's school career successive teachers annually entered averages of achievement grades in reading, arithmetic, handwriting, spelling, physical education, and other subjects. Teachers also recorded their estimates of such personal characteristics as level of effort, reliability, cooperation, leadership, and initiative as well as information on attendance and tardiness. In addition, space was available for teachers, principals, and secretaries to enter unstructured comments about the pupils as they saw fit. Hence the cards frequently included such statements as "family always fighting" or "this family very poor," as well as remarks about individual children such as "Jimmy is a nuisance and has to be watched all the time," or "Susie is a lovely child and a joy to have in the classroom."

Although the search through the record cards was initially undertaken as a means of locating former students, an unexpected phenomenon consumed our attention for some time. We were struck by the frequent and substantial variation in the IQ scores listed for the same children over time. The most extreme case was that of a girl whose recorded IQ scores differed by thirty-three points. In the third grade, her IQ was listed as 93, and in the sixth grade it was 126.

According to some psychologists, IQ is a stable characteristic and changes very little beyond school-entry age (Bloom, 1964). We are not prepared to engage in the argument over whether intelligence changes, and, indeed, we abandoned this approach because further unexpected findings emerged during the analysis. There is no question, however, that many of these children's IQs, as listed many years ago on their permanent record cards, show substantial variations over three-year periods.

At this stage of our inquiry, Rosenthal and Jacobson's *Pygmalion in the Classroom* (1968) was leading many educators to believe that teacher expectations could systematically alter pupils' IQs. Nonetheless skeptics such as Thorndike (1968, 1969) offered equally plausible alternative hypotheses. Some argued, for example, that the observed differences in IQ resulted simply from measurement error. To obtain a quick test of this hypothesis, we drew at random fifty permanent record cards for boys and fifty for girls and analyzed the association betwen sex and IQ change by means of a chi-square test. This analysis showed that girls were about as likely to increase in IQ as they were to decrease but that boys were twice as likely to decrease as they were to increase. The probability of chance occurrence of the observed pattern of difference was less than .02, and we interpreted this finding as statistically significant. If the observed IQ changes had indeed resulted from error of measurement, the error was random for girls but systematic for boys. This analysis supported the idea that higher teacher expectations for the academic achievement of girls as opposed to that of boys were part of the reason for the observed scores. The difference in patterns of IQ changes by gender encouraged us to develop a more extensive research project aimed at fully exploiting the available data and gathering more information to study the phenomenon.

Our earliest plans were more grandiose than proved feasible. We envisioned three stages of research: the first to analyze the data already available in the permanent record cards, the second to supplement this analysis by additional data about the subjects' adult status, and the third to observe teachers working with students of various social-class levels. The three studies were to be done in sequence such that the results from one study phase would determine the design for the next phase. This report, however, only describes some of the findings of the first two phases; the third phase was never begun.

The First Stage of the Study

The first stage of the study, which is based entirely on data from the school records, is presented in brief form. Detailed descriptions of findings will be given only for those parts that led directly to the change in plans for the later analysis.

The general idea behind the early stage of research was that, in some way, the child's evolving academic self-concept arose out of background experiences, personal attributes, and school experiences. Thus, a negative academic self-concept could result from an impoverished background or from exposure to teachers whose expectations for pupil achievement were low. Negative self-concept and associated low effort would result in generally low levels of motivation and relatively poor learning. This would lead to low IQ-test performance in the third grade and even poorer results in the sixth; in sixth grade, the ability to read the IQ test items accurately and quickly, while essential, would most likely be depressed in students suffering from cumulative deficit.

We began with five sets of hypotheses. The first four related IQ change to background, school experience, personal characteristics, and school achievement. The last was a quasi-longitudinal analysis examining IQ change as a predictor of teacher comments. Testing these hypotheses involved searching for statistically significant associations between IQ change and the variables mentioned above. Because our data were not gathered experimentally, the reader is cautioned against drawing causal inferences.

As suggested above, our first set of hypotheses dealt specifically with background factors associated with subsequent IQ change. These factors included gender; home stability, defined by the incidence of record-card entries such as "parents always fighting" or other indications of family instability; completeness of family, defined by the presence or absence of one or both natural parents in the home; and geographic mobility, indicated by the frequency of change-of-address notations on the record cards. Our guiding hypothesis was that unfavorable family background and male gender were associated with IQ decrease.

The results of the analyses were mixed but generally supported our initial hypothesis. Boys were found more likely to experience IQ decrease than girls[2] and high geographic mobility was associated with IQ decrease. Home stability was not significantly related to IQ change. Completeness of family, however, was found to be significantly related but not in the simple pattern we had expected. Rather, we

[2] The .05 level of significance is used unless otherwise noted.

TABLE 1
Percentage of Pupils Experiencing IQ Change, by First-Grade Teacher

	First-Grade Teacher			
IQ Change	*Miss A*	*Miss B*	*Miss C*	*Others*
Increase	36%	31%	12%	34%
No change	31%	46%	48%	36%
Decrease	33%	23%	40%	30%
Total	100%	100%	100%	100%
(*n*)	(78)	(69)	(40)	(120)

$x^2 = 10.7$; $df = 4$; $p < .05$; $N = 307$

noted a tendency toward stable IQ in children from families with both parents, whereas IQ increase most often occurred in children with only one parent at home. IQ decrease was the modal pattern for orphans and foster children. This latter set of relationships concerning completeness of families is significant at the .01 level of confidence. We concluded that these findings supported the hypothesis that background and gender do influence IQ change.

In addition to varying family backgrounds, children can be subjected to different school experiences. If teacher expectations for pupil performance can be manipulated, as Rosenthal and Jacobson (1968) have asserted, then one might also postulate that different teachers, because of their individual outlooks, may have different levels of expectations for children. And if evolving pupil self-concept has an impact on IQ-test performance, then experiences such as failing and repeating grades or receiving negative teacher comments should also have a similar bearing.

Our second general hypothesis was that IQ decrease is associated with indications of low academic ability, such as experiencing a year with a first-grade teacher who held low expectations for pupil performance, receiving negative teacher comments, and failing or having to repeat grades. The analysis showed the predicted relationships between failure and IQ change ($p < .001$) and teacher comments[8] and IQ change ($p < .02$). These results lent further support to our general hypothesis. The relationship between first-grade teacher and IQ change is presented in detail, because it shows that the subsequent pattern of IQ change does significantly vary with first-grade teacher, and because this finding led to the change in the design of our analysis. The relevant data are presented in Table 1.

Table 1 presents the percentages of pupils experiencing IQ change, categorized by teacher and by direction of change. Miss A, Miss B, and Miss C were teachers who had worked at the school for many years, whereas "Others" are a collection of

[8] It is discouraging to find that teachers were about twice as likely to warn future teachers of the shortcomings of their pupils as they were to encourage a positive mindset. If this is an indicator of general attitudes of teachers toward pupils in this school, such negative teacher outlooks might help to account for the diminishing level of motivation over the years observed in this study. An interesting parallel observation is Rist's (1970) comment about negative pupil characteristics sought out during the initial interview with parents prior to the entry of children to school.

individuals who stayed for short periods and then moved on. This table shows, for example, that while 36 percent of Miss A's pupils subsequently experienced IQ increase, only one-third that proportion—12 percent—of Miss C's children did. The differences in these patterns are statistically significant at the .05 level. We will return to a more refined analysis of this phenomenon after we have further examined the idea of IQ change as a label influencing teacher attitudes toward pupils and expectations for pupils' success in school.

Our idea of the relationship between teacher influence and IQ change involves the notion of the self-concept. Unfortunately, the permanent record cards provided no direct measure of self-concept. We were convinced, however, that some of the personal characteristics estimated by the teachers could be assumed to be associated with pupils' academic self-concept. Therefore, we focused on effort and on leadership and initiative as indicators, or proxy measures, of self-concept. We hypothesized that high scores on these variables would be associated with IQ increase. The analysis strongly supported this hypothesis, the level of significance being .01 for leadership and initiative and .001 for effort.

The analyses completed at this point supported the association between IQ change and background, first-grade teacher, and self-concept. The next general hypothesis, an extension of this analysis, is that high achievement grades in school are likely to be associated with IQ increase. We examined this hypothesis in two ways: first, reading achievement in the first grade was used as a predictor of later IQ change; then, school achievement grades averaged over the whole school career were examined in relation to IQ change. Both analyses supported the hypothesis. For example, 35 percent of pupils whose first-grade reading marks were very good or excellent experienced IQ increase, whereas only 18 percent of pupils with fair or unsatisfactory reading grades subsequently experienced IQ increase ($p < .001$). Also, grades averaged over the whole school career in reading, language, arithmetic, French, history, and even art were significantly associated with IQ change (each $p < .01$).

The fifth hypothesis, which was analyzed in a quasi-longitudinal pattern, was that teacher comments, negative or positive, are directly related to IQ change. The question behind this analysis is whether or not teachers' attitudes toward pupils are influenced by labels such as IQ or by patterns of label changes, such as IQ change. This question is valid whether or not the child's intelligence was really changing, because teachers' expectations for pupil achievement could be influenced by the scores entered in the child's records.

Because the results of this analysis changed the direction of our research, we will present them here in some detail. Since teacher comments were signed or initialed, it was possible to determine at which grade the comments were made. It was thus possible to determine the negative or positive quality of teacher comments *following* IQ change. The reader will recall that the first IQ test took place in the third grade and the second during the sixth grade. The sixth-grade teachers were required to enter the results of the sixth-grade IQ tests on the permanent record cards by placing them just below the third-grade test results. Thus, while entering the second of two results, the sixth-grade teacher would become aware of whether IQ

had changed or not and, if so, in what direction and by how much. Before the end of that school year, the same teacher would enter comments on the permanent record cards. In the following September, the seventh-grade teacher received the permanent record cards for these pupils and may well have noted pupil IQ-change patterns prior to writing her own comments on the cards. We therefore had a situation that could be analyzed to determine if the subsequent comments listed by the teachers were in the same pattern as the perceived IQ changes.

Our analysis, illustrated in Table 2, showed, for example, that sixth- and seventh-grade teachers were more likely to write positive comments for children whose IQ had increased than for children whose IQ had decreased or remained the same. Thus, regardless of how IQ changes came about, it seems likely that, once the pattern of change had been established, subsequent teachers may have reinforced this pattern. This is a tentative conclusion resulting from scanty evidence. Other explanations are possible; ours, however, is consistent with the theory of the self-fulfilling prophecy proposed by Merton (1957) and illustrated in school settings by Rist (1970).

We returned to a more refined analysis of the relationship between first-grade teacher and IQ change because it was here that we were examining the effects of IQ change, considered as a label, on teachers. We intended to test the hypothesis that teachers' attitudes toward children can be affected by different labels such as IQ and gender. Table 1 reported the relationships between first-grade teachers and IQ-change patterns of the children. In this analysis we restricted our interest to the three long-service teachers because we wanted to study the effects of labels on individual teachers, not on categories of teachers.

It was possible, by identifying the gender of the pupils of these three teachers, to consider gender as a label and use it as a factor in our analysis. We were quite certain that, if there was bias in assigning children to these teachers that was systematically related to the academic potential of the children, gender was not a significant element of that bias. In other words, there was no reason to suspect that any first-grade teacher was given smart girls and dull boys, no matter how estimates of the academic potential of entering first graders might have been assessed at that time. Furthermore, if a bias among the pupils of one teacher was not balanced by the

TABLE 2

Percentage of Teacher Comments Following Pupils' IQ Change

	IQ Change		
Teacher Comments	*Decrease*	*No Change*	*Increase*
Positive	9%	8%	15%
None	70%	83%	77%
Negative	21%	9%	8%
Total	100%	100%	100%
(*n*)	(93)	(116)	(94)

$\chi^2 = 12.11$; $df = 4$; $p < .02$; $N = 303$

TABLE 3
Percentage of Pupils Experiencing IQ Change, by Sex of Pupil and First-Grade Teacher

	Miss A		*Miss B*		*Miss C*	
IQ Change	*Boys*	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*
Increase	33%	38%	18%	39%	12%	13%
No change	33%	28%	43%	49%	41%	52%
Decrease	33%	33%	38%	13%	47%	35%
Total[a]	99%	99%	99%	101%	100%	100%
(*n*)	(39)	(39)	(28)	(41)	(17)	(23)

Miss A: $\chi^2 = 1.0$; $df = 2$; $p > .70$;
Miss B: $\chi^2 = 8.5$; $df = 2$; $p < .02$;
Miss C: $\chi^2 = 0.5$; $df = 2$; $p > .80$;
$N = 287$

[a]Some columns do not sum to 100% because of rounding off.

opposite bias among the pupils of the other teachers, it is likely that there was no differential allocation of children by ability and by gender. Therefore, we examined the subsequent IQ-change patterns of the pupils of Miss A, Miss B, and Miss C separately for each gender, without any other hypothesis than that we expected differences to emerge if the teachers had different attitudes about boys as compared to girls.

Table 3 shows that among Miss A's pupils, the boys' IQs seemed somewhat less likely to increase than the girls', but the differences are not statistically significant. In Miss C's class, although her boys' IQs seemed more likely to decrease than did the girls', the difference is still not statistically significant. In effect, it appears that to be assigned to Miss C was unfortunate no matter what one's gender. The very curious pattern for Miss B lent support to the hypothesis that a teacher may have different reactions to children with different labels. As Table 3 shows, the IQs of boys from Miss B's class were about twice as likely to decrease as they were to increase, and the IQs of girls in the same class were three times as likely to increase as to decrease ($p < .02$).

Miss B was the author's first-grade teacher when he was a pupil at Ray School. While the author has very few memories of Miss B's classroom, others who were in her class recalled that Miss B had a differential seating policy for boys and girls. In the Ray School, all the seats were screwed to the floor in six rows of eight seats each. Miss B's policy was to seat the girls in the first four seats of each row and to place the boys in the last four seats. Teachers often assign seats in order to achieve classroom control; children sitting in the front rows work more attentively and feel smarter and better liked by the teacher than those in the back rows (Schwebel & Cherlin, 1972). But we concluded further that Miss B's differential seating pattern indicated different expectations of boys' and girls' school abilities. Since girls were seated in the front rows, it is probable that she had higher expectations for the academic achievement of girls than of boys. The bias evidenced by her seating arrangement may also have been expressed in the nature and degree of her inter-

action, as well as the quality of her teaching, which may have consistently bestowed advantages on girls and disadvantages on boys.

It is now many years too late to make any detailed study of Miss B and her behavior toward these children. The only conclusion we wish to draw from this analysis is that teachers may respond to particular labels in different ways. Furthermore, these differing responses may lead to different—and academically significant—treatment of certain children, which, in turn, may lead to different levels of success in schooling and influence future capacities, personal characteristics, and achievements.

The foregoing analyses were based on data available in school files for many years. The questions raised, however, could not be answered with only the data on the permanent record cards. We decided, therefore, to move to the next stage of this study—the interview stage. Before describing the design of the interviews, we discuss in some detail the questions influencing the design of this next stage of our research.

The Second Stage of the Study

The Problem and Theoretical Background

We have already implied that some pupils in a school may have experiences that enable them to adapt to and cope with both school and society, while the experiences of other pupils prevent them from effectively meeting increasing demands for achievement. These observations are based on the analysis of data available on the permanent record card of each pupil. The information, however, came to an end after the seventh grade, when students were about thirteen or fourteen years old. Naturally, we wondered what had happened to school children following graduation from Ray School. Were they caught up in the so-called welfare cycle? Had there been any long-term effects of schooling on this cycle? Did the children of families on welfare end up on welfare themselves? Could it be shown that certain kinds of experiences and achievements in elementary school would lead to upward mobility? The answers to such questions would require data to be collected by interviewing some of the former pupils.

Questions to be asked always depend on a theory to be examined; we therefore give a brief description of the assumptions that shaped our inquiries. Our work was based on the notion that—other things being equal—pupils who are expected to do well demonstrate higher achievement than those who are expected to do poorly. Furthermore, if pupils get a good start in school, they will benefit not only throughout their school careers but throughout their adult lives as well.

The basic model guiding our work (Brookover & Gottlieb, 1964) can be presented in terms of a "significant other"—a socializing agent interacting with a subject being socialized. If the significant other has positive expectations, his or her behavior during social interaction conveys those expectations to the subject. Since the subject regards the significant other as knowledgeable, she or he accepts the implied positive judgment, thus beginning the development of a positive self-

concept. Having been led to expect success, the subject is motivated to try hard. Among comparable pupils, those who work harder usually achieve more. In school settings, the significant other notes and rewards achievement in a variety of ways, and these positive evaluations, as well as other, more objective indicators of success, reinforce the already positive self-concept. Continuing effort becomes habitual and a tendency to succeed becomes part of the subject's personal reputation—another sort of label. Colloquially, the subject is regarded as a "winner." This reputation, as well as more objective indicators such as test results and reports of academic standing, influence, in turn, the expectations of subsequent significant others in related fields of endeavor. In addition, the self-confidence of the subject is apparent in classroom behavior and leads to more positive expectations in other social interactions. The original positive expectations of the first significant other have thus been justified, and the self-fulfilling prophecy has come true (Merton, 1957).

The model presented above raises a number of questions. One question that is basic to our research is: where do the expectations of the significant other originate? To answer it, one can focus on the personality and general outlook of the significant other, on the characteristics of the subject, or on some combination of both. Different foci will give rise to different research designs. Rist (1970) chose to focus on the subjects, whereas our research was more concerned with assumed differences among the significant others.

In his research, Rist followed a single class through kindergarten and first and second grades. He noted that the teachers treated each child differently and that this differential treatment corresponded to differences in the social background of the subjects. Rist found that the first teacher to whom the children in his study were assigned expected relatively disadvantaged children to do poorly. He drew these conclusions not only by observing the teacher's behavior but also from the teacher's own statements. The differential treatment of children produced the development of behavior and achievement patterns that the teacher expected.

In the work we describe here, we took a slightly different approach, which led to the conclusion that children of substantially similar backgrounds who were assigned to different teachers exhibited different levels of success. We believe that some teachers are more positive in their attitude toward children and are, perhaps, better teachers. If children are fortunate enough to begin their schooling with an optimistic teacher who expects them to do well and who teaches them the basic skills needed for further academic success, they are likely to perform better than those exposed to a teacher who conveys a discouraging, self-defeating outlook.

The Research Design and the Interview Schedule

Since we focused on IQ change as our major dependent variable in the first stage of the study and because we chose our sampling method at that time, we decided to use IQ change as an element in the choice of the subjects. Our data were drawn from approximately a twenty-five-year time span and were designed to link school experience with later adult achievement. Random sampling of students was possible, but actually locating the selected persons who are now adults was clearly

impossible. We calculated that our resources would allow us to interview sixty subjects but were concerned that a loss in sample size would seriously jeopardize the possibility of significant findings. To offset this danger, we decided to choose subjects concentrated at the extremes and the middle of the IQ-change distribution. We identified one hundred subjects of which we were finally able to find and interview sixty. (One of these turned out to be a case of mistaken identity and was dropped from subsequent analyses.) The resulting sample of thirty females and thirty males was divided into three groups of ten females each and three groups of ten males each who had experienced either an IQ increase, an IQ decrease, or virtually no change in IQ as recorded in the third and sixth grades.

At that time, our theoretical model assumed, in part, that apparent IQ change influenced teachers' expectations for their pupils. We believed that changes of four points or less would have been dismissed as falling within the standard error of the tests; we expected, however, that teachers probably would have been influenced by changes as large as ten points or more, since changes of such magnitude approach or exceed one standard deviation on a normal IQ distribution.

Unfortunately, despite these criteria, we were not able to obtain ideal IQ-change characteristics for all of the subjects interviewed. Although we had no difficulty obtaining twenty cases who had experienced four points or less change, we had to include seven cases with changes smaller than ten points, in order to obtain our interview sample of sixty cases. Table 4 summarizes the IQ-change patterns of the final sample interviewed, and shows large contrasts among the subjects. The sixty individuals interviewed may be classified into three groups of twenty each with very different IQ-change patterns. For example, the figures for boys in the decrease row show that the boy with the smallest decrease had a change of seven points from the third to the sixth grade, whereas the boy with the largest decrease had a change of twenty-one points.

As stated previously, our design of the interview sample was greatly influenced by Rosenthal and Jacobson's (1968) notion of the "Pygmalion effect" on IQ change. We had already embarked on the interview schedule, modeled on the design illustrated in Table 4, when the correlation of first-grade teacher effect with pupil gender (Table 3) shifted our interest from IQ change to teacher effect. This shift in focus was strengthened by the criticisms of the work of Rosenthal and Jacobson and by the appearance of Rist's article, to which we have already referred. Because of limitations in time and resources and the career demands of the researcher, we

TABLE 4
IQ-Change Pattern of the Thirty Boys and Thirty Girls in the Interview Sample

	Range of Change Magnitudes		*Mean Change Magnitude*		
IQ Change	*Boys*	*Girls*	*Boys*	*Girls*	*(n)*
Increase	8 to 24	8 to 33	13.0	14.3	(20)
No change	1 to 4	1 to 4	2.6	2.5	(20)
Decrease	7 to 21	9 to 21	14.5	15.9	(20)

moved ahead to the revised second stage without a revised sample. We will say more later about the shortcomings of this sample for our purposes as they emerged.

Having decided upon the size and type of sample, we had to address the problem of the substance and design of the interview schedule. Our subjects were said to be disadvantaged in part because they were living under conditions of relative poverty. Inspired to some extent by the notion of the culture of poverty, we developed the following general hypothesis: in cases where schooling inculcates children with middle-class values and attitudes, upward social mobility is likely to occur; but in cases where schooling fails to achieve this effect, individuals continue in adulthood to exhibit values, attitudes, and personal characteristics associated with the culture of poverty. This general hypothesis generated a number of more specific hypotheses and questions to be examined with interview data. Only a very small part of all these efforts, however, is reported here. The data used here bear mainly on two variables—first-grade teacher and adult status. Pupils were asked to evaluate the effectiveness of their former teachers, to grade them for effort, and to make unstructured comments about their overall performance. At the same time the interviewer was requested to make observations and inquire about the social status of the subjects.

We intended to study the social mobility of disadvantaged children by using variables from the permanent record cards (and recalled by adults from their childhood) as predictors of adult status. Questions were designed to extract indicators of adult status, to be combined for use as a single dependent variable. The achievement of relatively high adult status would in this sense become the operational definition of upward social mobility. We therefore devised a number of questions concerning occupational status and occupational history in order to obtain a list of positions the subjects had held since entering the work world. Moreover, in order to obtain other indicators of adult status, we asked questions about the type of dwelling, amount of rent paid, the number of rooms, the state of repair, cleanliness and orderliness, and the condition of the furnishings.

Occupational prestige became an important component of our analysis of adult status. Our initial attempt to assign values to the subjects' occupations using the Blishen (1958) scale was ultimately unsuccessful. The range of occupations from such a uniformly low socioeconomic background was simply not broad enough to be captured by that method. We also came to the conclusion that we were really not interested in comparing our subjects to any external standard but merely to get maximum discrimination among them. Thus, we decided to list each subject's occupation on two identical sets of small file cards; occupations ranged from "professor of sociology in a junior college" to "unemployed, has never worked." Next, two experts in job classification each sorted one set of the cards into ranks, from highest to lowest occupational prestige. The judges worked independently, arranging the subjects into ranks from one to fifteen. They were then instructed to collapse the fifteen ranks into five ranks of any convenient size. After this procedure was completed the two sets were compared. Of the sixty cases, judges were in agreement on forty-two; there were twelve cases in which subjects were in different but adjacent categories and six where the differences were larger

TABLE 5
Components of Adult Status and Their Values

	Values Added to Arrive at Total Adult Status Score						
Component	*1*	*2*	*3*	*4*	*5*	*6*	*7*
Highest grade	VII or lower	VIII	IX	X	XI	XII or higher	
Rent paid	$55 or less	$56–$79	$80–$89	$90–$99	$100–$119	$120–$134	$135 or more
Type of housing	flat	multiple dwelling	apart-ment	house			
Repair of housing	poor	fair	good	very good	excellent		
Personal appear-ance	poor	fair	good	very good	excellent		
Occupa-tional level	low	moderately low	moderate	moderately high	high		

than one category in the hierarchy. Despite these measures, we decided arbitrarily on the final ranking of the twelve cases in different but adjacent categories, and for those with larger differences the in-between rank was assigned.[4]

Other variables thought to be related to adult status were included in the interview schedule; thus, subjects were asked to indicate the age and grade when they had left school. In addition, our interviewer was required to comment on the personal appearance of the subject being interviewed. Recognizing the subjective quality of these estimates, we took steps to assure greater objectivity. The interviewer was first asked not only to evaluate the general appearance but also to note such specific things as the style of dress, grooming habits, and body posture. Precautions were taken to make sure that the interviewer had none of the information on the school record cards and was ignorant of the hypotheses about to be tested.

Table 5 illustrates the components of adult status and the values used to determine an adult-status score. Two examples will help explain how to read Table 5: having completed grade twelve or higher contributed six points to one's adult-status score whereas having completed only eighth grade contributed two points. The second-lowest occupational prestige rank—"moderately low" in the table—made a contribution of two points.

No attempt was made to factor analyze these indices and weight them for the

[4] The 70 percent level of agreement is a low estimate of interrater reliability because it considers all failures in agreement as equal, even though the subjects may be in adjacent categories. We also calculated the rank-order correlations of the judges' ratings using their original fifteen-category classifications. The Spearman rho, corrected for ties, was equal to .82 (see Siegel, 1966, for discussion).

development of an adult-status scale because such sophisticated treatment of relatively crude data might create more confidence in the results than could be justified by the original accuracy of the component data. The intercorrelations among these six component variables used summatively to create the variable adult status range between .50 and .80. To arrive at adult-status rankings, each subject's scores were summed along the six indicator variables. This aggregation was predicated on the assumption that the more indicators considered in deriving adult-status rankings, the less likely it would be that some atypical scores on one of the variables would result in an inappropriate rank for an individual. The resulting total index had a possible range from six to thirty-two points and was cut into nine approximately equally spaced ordinal ranks for the adult-status score.

The distribution of adult status is presented in Table 6. In some of the analyses presented later, the adult-status ranks were further collapsed to produce only three ranks with twenty-two subjects in the high category, twenty subjects in the middle category, and seventeen subjects in the low category. Scale 1, with nine categories, is presented in the left section of Table 6; Scale 2, a three-category scale, is presented on the right.

Findings from the Second Stage of the Study

The data obtained from interviews were added to information already available in the permanent record cards, and the overall analysis was conducted along three different lines of inquiry. At first, simple contingency tables were developed to show the relationships between variables (sometimes for subclasses of subjects) and significance was determined by the application of the chi-square test. Following that, analysis of variance and the *F*-test for significance were used to compare means.

At the outset of this article, we reported that certain trends in educational research and unanticipated results deflected our interests and changed our focus

TABLE 6
Distribution of Subjects on Adult-Status Scales

Level	*Scale 1*	*Number of Cases*	*Scale 2*	*Number of Cases*
Highest	9	6		
	8	6	3	22
	7	10		
	6	5		
Medium	5	7	2	20
	4	8		
	3	6		
	2	6	1	17
Lowest	1	5		
Total		59		59

during the study. The result of a simple cross-sectional analysis of the variables "first-grade teacher" and "adult status" precipitated such a change in focus. If we had known that the emphasis of our study would shift from IQ change to the effect of first-grade teacher, our sample design for the interviews would have been different. We would have included equal numbers of Miss A's former pupils and of the former pupils of other first-grade teachers, all selected at random. Because it was too late to start over, we decided to continue, knowing that we now had a less-than-appropriate sample. In describing the results of this stage of our work, we shall, with one exception, present a chronological narrative because we believe that the process we are analyzing is developmental—one in which the sequence of events helps to determine the ultimate outcome.

Before we begin, however, the one exception commands our immediate attention: the remarkable relationship we observed when we cross-tabulated adult status with first-grade teacher and performed an analysis of variance (Table 7A).

Analysis of Table 7A shows a marked difference between the adult status of the individuals we interviewed who as children had been assigned to Miss A and those who had been assigned to Miss B, Miss C, and other teachers. There was remarkably little difference between the mean adult status of those assigned to Miss B, Miss C, and Others; and, indeed, these differences are clearly not significant. For greater simplicity of presentation, we have not shown the patterns by gender because we found this variable did not affect adult status. Miss A's pupils, boys and girls, did well; by comparison, pupils of other teachers, regardless of gender, did poorly.

We have combined in Table 7B the pupils of Miss B, Miss C, and Others into a single category called "Others" to simplify the table and to minimize problems of analysis arising from small numbers in cells of contingency tables. Table 7B shows that Miss A's superiority as a teacher, inferred from her reputation, to be described later, seems to have some relationship to the subsequent performance of her pupils. For example, two-thirds of her pupils achieved the highest adult status, compared to less than half that proportion for the former pupils of the other first-grade teachers. None of Miss A's pupils was in the lowest status category, but more than a third of other respondents had this unhappy experience.

TABLE 7A
Adult Status, by First-Grade Teacher

	First-Grade Teacher			
Adult Status	*Miss A*	*Miss B*	*Miss C*	*Others*
High	64%	31%	10%	39%
Medium	36%	38%	45%	22%
Low	0%	31%	45%	39%
Total	100%	100%	100%	100%
(*n*)	(14)	(16)	(11)	(18)
Mean adult status	7.0	4.8	4.3	4.6

$F = 3.49$; $df = 3, 52$; $p < .02$; $N = 59$.

TABLE 7B
Adult Status, by First-Grade Teacher

Adult Status	*First-Grade Teacher*	
	Miss A	*All Others*
High	64%	29%
Medium	36%	33%
Low	0%	38%
Total	100%	100%
(*n*)	(14)	(45)
Mean adult status	7.0	4.6

$F = 12.03$; $df = 1, 57$; $p < .001$; $N = 59$

These findings are so dramatic that one might suspect the results are spurious. Indeed, colleagues challenged the results and proposed alternative explanations. These challenges were, by and large, concerned with two questions. First, were there any factors operating prior to schooling that could influence the school career and adult status of the subjects? In other words, did Miss A receive from the very beginning pupils of higher quality than did other teachers? Second, could we demonstrate how first-grade teachers actually influenced their pupils' later achievements? Was there a process of influence that we could identify and describe?

Rist, responding to an earlier version of our paper, suggested that the results might be spurious on the grounds that Miss A had perhaps received more than her fair share of children from relatively good social backgrounds. On the face of it this possibility seemed unlikely: the school attempted to assign children equally to all teachers; no objective screening devices such as IQ tests were employed for the assignment of children to any teacher on the basis of ability; and the fourteen former pupils of Miss A in our sample did not come from a single cohort but from classes that spanned an eleven-year period.

We felt, however, that Rist might have a point. Referring to the permanent record cards and the interview schedules[5] for background information such as father's occupation[6] and completeness of the family, we noted some differences sug-

[5] It should be noted that the information given in the interviews generally agreed with the information on the permanent record cards. This gives us some confidence in the memory of the adults interviewed, as well as in the accuracy of the information on the permanent record cards.

[6] The occupation of the father was indicated both on the permanent record card and in reply to a question in the interview of adult subjects. Occupations from these sources were compared for each father ($N = 48$) and were ranked separately by each co-author, after consultation with an expert from the personnel office of a large corporation. Two scales were thus constructed—a four-degree scale and a five-degree scale. The combined scores from both scales were reduced to a six-degree scale. The absence of the father and the occupation of the mother were taken into consideration in the scale. The range of occupations and economic situations is rather narrow. Still, distinctions could be made as shown below:

Rank	Occupations
6	Porter or waiter on railways, coremaker, blacksmith, plumber, carman.

TABLE 8

Relationship of Miss A to Adult Status, with Effect of Background Variables Removed

Variables in the Equation	Cumulative Variance Explained (R^2)	Standardized Regression Coefficient (β)	Incremental F
Number of children	.139	−0.33	7.5*
Father's occupational status	.220	0.23	3.1
Family ever on welfare	.228	0.11	0.7
Miss A	.331	0.34	8.3*

*$p < .01$.

gesting that Miss A had been slightly favored. For example, father's occupational status was positively correlated with being in Miss A's class ($r = .20$) and positively correlated to adult status ($r = .24$). Being from large families was negatively related to adult status ($r = -.37$), and Miss A had slightly less than her share of children from large families ($r = -.13$).

To unravel these differences, we performed a modified stepwise-multiple-regression analysis relating father's occupational status, welfare status, number of children in the family, and membership in Miss A's class to adult status as the dependent variable. We directed the computer program to enter Miss A as the last variable, so that we could calculate the increment to explained variance solely attributable to her.

Table 8, which presents the results of this regression analysis, shows that the background variables explained nearly 23 percent of the variance in adult status. Miss A added 10 percent of explained variation, which is significant at the .01 level (see Cohen, 1968). This test provided a conservative estimate of Miss A's effect, and finding a high level of significance with a small sample size gave strong support to the likelihood that she had influence on adult status over and above the effects of background variables. We could not completely eliminate the possibility of spuriousness, because we could not control all possible background variables. However, the background variables we were able to control are powerful ones; therefore, the probability that the relationship of Miss A to adult status was spur-

5	Salesman, typesetter, postal clerk, clerk, dairy foreman, taxi driver.
4	Checker, milliner, commercial artist, carpenter.
3	Mechanic, cook, truck driver, laborer.
2	Laborer, cleaner, janitor, truck driver, construction worker, butcher, machine operator, painter-laborer.
1	Father chronically ill, unemployed, absent, or on welfare; mother dishwasher, houseworker, switchboard operator, laborer, waitress, on welfare.

In arriving at these ranks, consideration was given to whether or not the laborer, for example, was a long-term employee of a company that had enrolled him in a pension plan, or whether he simply drifted from job to job as a laborer without any chance of accumulating fringe-benefits or job security.

ious is greatly reduced. Our task now is to explain how her apparent influence was actually achieved and to suggest, if possible, the process by which the assumed pattern of Miss A's positive expectations became a self-fulfilling prophecy.

First, however, we should discuss Miss A. Since she was at the late stages of a terminal illness when we began the interviews, we could not observe her in the classroom or ask her to explain her effects on children. We can offer a description of Miss A's remarkable reputation as a teacher. Not only did Miss A have a good reputation as a teacher in Ray School, but she was still remembered by the individuals we interviewed about a quarter of a century after they had been her students. During the interviews, the subjects were encouraged to recall and evaluate each of their elementary-school teachers. These former students were instructed to use the same scheme that had been applied to them when they were in school; hence, respondents graded their old teachers E, VG, G, F, or U (excellent, very good, good, fair, or unsatisfactory) for achievement, and A, B, C, or D for effort. Many of the subjects could not remember the names of their teachers, and sometimes their memories were actually wrong, as we discovered when comparing their recollections with information on the permanent record cards. Despite the general difficulty of remembering every teacher, not a single subject who had been in Miss A's class failed to recall that fact correctly. Of those who had *not* been Miss A's pupils in first grade, 31 percent had no recollection of who their teacher had been, fewer than half identified their first-grade teacher correctly, and four subjects incorrectly named Miss A as their teacher. Memory seems to have been influenced by wishful thinking!

There are other indications of Miss A's reputation as a good teacher. Of those who had been members of Miss A's classes, more than three-quarters rated her very good or excellent as a teacher, whereas of those assigned to other first-grade teachers, fewer than one-third described them as very good or excellent. The grades for effort are similar: 71 percent of the former pupils of Miss A rated her effort A, whereas only 25 percent of the pupils who had been in the classes of other first-grade teachers rated them A.

Reputations of teachers do not in themselves account for the performance of their pupils, but it is worth noting that after hearing an informal presentation of one of our interim reports, Robert Collins, a staff writer of *Reader's Digest,* became interested in Miss A as a topic for an article. We gave him leads on former pupils and colleagues of Miss A, and he conducted a number of interviews, not only with former pupils but also with teachers and principals who had worked with her. The article corroborated our findings concerning Miss A's reputation. Collins (1976) reported that Miss A had taught for thirty-four years in Ray School and that "she kept control by sheer force of personality and her obvious affection for the children, never needing to lose her temper or resort to physical restraint" (p. 144).

It was said of Miss A's teaching that "it did not matter what background or abilities the beginning pupil had; there was no way that the pupil was not going to read by the end of grade one" (p. 144). One informant reported that Miss A left her pupils with a "profound impression of the importance of schooling, and how one should stick to it" and that "she gave extra hours to the children who

were slow learners" (p. 144). When children forgot their lunches, she would give them some of her own, and she invariably stayed after hours to help children. Not only did her pupils remember her, but she apparently could remember each former pupil by name even after an interval of twenty years. She adjusted to new math and reading methods, but her secret for success was summarized by a former colleague this way: "How did she teach? With a lot of love!" (p. 144). One would add, with a lot of confidence in children and hard work.

Academic Achievement

If it can be shown that, on the average, Miss A's pupils reached higher levels of academic achievement than children who were placed with other teachers, we would have the beginnings of an explanation for their subsequent higher socioeconomic achievement as adults. Since teachers evaluated the achievement of pupils in each subject, recorded the grades on the pupils' permanent record cards, and then averaged all the academic grades into a mark called "General Standing," we used this average as an indicator of Miss A's effectiveness as a teacher. Because analyses of these data showed that achievement in virtually every elementary-school subject that has any academic content is highly correlated with achievement in reading (as we noted in our earlier analysis of IQ change and academic achievement), this average of all school subjects signified, in large part, the ability to read. Not only was this true in first grade but it also held for subjects such as geography and history in the higher grades. These high correlations with reading undoubtedly resulted in large measure from the traditional method of teaching

TABLE 9
Achievement in First Grade, by First-Grade Teacher[a]

Achievement in First Grade	*First-Grade Teacher*	
	Miss A	*Others*
High	64%	28%
Medium	29%	44%
Low	7%	28%
Total	100%	100%
(*n*)[b]	(14)	(39)
Mean achievement[c]	6.5	4.9

$F = 6.28$; $df = 151$; $p < .02$; $N = 53$

[a]A similar pattern occurs for Grade I Reading. Miss A's pupils appear to have higher grades than those of others; however, the difference is not statistically significant at the .05 level ($F = 3.79$, $df = 1.51$, $p < .057$).

[b]There were six missing cases for first-grade general standing.

[c]The means presented in this and following tables were calculated on the following basis: any pupil with excellent was scored 9; very good plus, 8; very good, 7; good plus, 6; good, 5; fair plus, 4; fair, 3; fair minus, 2; and unsatisfactory, 1. High achievement includes excellent and very good. Low achievement includes fair and unsatisfactory. Medium achievement includes good plus, good, fair plus.

TABLE 10
Pupils' Second-Grade Achievement, by First-Grade Teacher

Achievement in Second Grade	*First-Grade Teacher*	
	Miss A	*Others*
High	64%	28%
Medium	29%	36%
Low	7%	36%
Total	100%	100%
(*n*)	(14)	(39)
Mean achievement	6.5	4.9

$F = 6.54$; $df = 1, 51$; $p < .01$; $N = 53$

"content" subjects by assigning readings in textbooks and requiring oral or written responses to questions following such assignments.

Table 9 shows the percentage of high, medium, and low achievement scores and presents the mean achievement for Miss A's children compared to other teachers' pupils. The figures in Table 9 clearly indicate that Miss A's pupils achieved substantially higher grades at the end of the year than did pupils of other first-grade teachers. But this observation raises a serious question. Could these results simply indicate that Miss A was an "easy marker" compared to the other teachers? If so, the difference would not be meaningful, since it would result simply from a different standard of grading. One way to answer this question is to see what subsequent teachers thought about the performance of Miss A's pupils.

Table 10 is identical to Table 9 except that the achievement marks are those entered on the permanent record cards a year later by the second-grade teachers. Thus, we can compare the subsequent achievement level of Miss A's pupils to that of the pupils of other teachers without the possible bias that may be influencing the figures in Table 9. Each second-grade teacher would have received her fair share of children from Miss A's class and from the classes of other first-grade teachers. Table 10 gives the same impression as Table 9. Since the marks of Miss A's pupils were essentially the same in the second grade, we conclude that the apparently higher achievement of her pupils as shown in Table 9 could not have resulted from her being an easy marker. Based on an analysis of variance, these differences in achievement were found to be statistically significant ($p < .02$ for Table 9, $p < .01$ for Table 10).

There still is room for skepticism; one may ask how long these differences in achievement persisted, and whether they were sufficiently enduring to explain the long-term effect on adult status suggested in Table 7A. The permanent record cards enabled us to analyze differences in mean achievement throughout the school career of children who were in Miss A's class and of other children. The figures in Table 11 show that the differences in mean achievement between Miss A's children and other teachers' pupils were maintained throughout the elementary-school career. Although the magnitude of the differences changed, there was not a single

TABLE 11
Mean Achievement in Each Year of Elementary School, by First-Grade Teacher

Mean Achievement	First-Grade Teacher	
	Miss A	Others
Grade I	6.5	4.9
Grade II	6.5	4.9
Grade III	5.9	5.1
Grade IV	5.5	4.8
Grade V	5.7	4.6
Grade VI	5.0	4.2
Grade VII	5.0	4.1

year in which the mean scores of Miss A's children were not higher than those of others. This result suggests that a good beginning in school will indeed have beneficial effects throughout one's schooling.

The reader will recall that our basic hypothesis dealt with the importance of a positive self-image. If children are lucky enough to have a first-grade teacher who has high expectations for their achievement, they will be more likely to develop positive academic self-concepts and to be more successful in school. Unfortunately, as we noted earlier, we had no way of measuring directly what the self-concept of a child was thirty years ago; however, some plausible ways of obtaining indirect indicators of self-concept were available. For one thing, children who have positive self-concepts in school would probably work harder than those who have negative feelings about their academic potential. The resulting expenditure of effort should produce achievement close to a pupil's maximum potential, and high achievement should, in turn, enhance the self-concept. We thought it reasonable, then, to use effort as a proxy measure for self-concept, since a more direct measure was not available.

Teachers reported to parents several times a year on a pupil's level of effort and summarized these grades on the permanent record cards. These marks for industry enabled us to compare the level of effort for Miss A's children with that

TABLE 12
Mean Effort in Each Year of Elementary School, by First-Grade Teacher

Mean Effort	First-Grade Teacher	
	Miss A	Others
Grade I	7.7	5.3
Grade II	7.3	6.2
Grade III	7.0	5.8
Grade IV	7.1	6.2
Grade V	6.7	6.6
Grade VI	6.8	6.2
Grade VII	7.0	6.2

TABLE 13
Mean Leadership and Initiative in Each Year of Elementary School, by First-Grade Teacher

	First-Grade Teacher	
Mean Leadership and Initiative	*Miss A*	*Others*
Grade I	5.4	4.5
Grade II	6.3	5.3
Grade III	6.5	5.1
Grade IV	5.8	5.1
Grade V	6.2	5.7
Grade VI	6.5	5.8
Grade VII	6.9	5.9

of others. Inspection of Table 12 shows that in every year pupils originally assigned to Miss A had higher mean levels of effort than other pupils; this may indicate a more positive self-concept. The permanent record cards contain other proxy measures for self-concept; for example, teachers also recorded estimates of leadership and initiative. Table 13 shows that the pattern for leadership and initiative resembled the pattern for effort. Although the magnitudes of differences varied, teachers' estimates of leadership and initiative for Miss A's pupils were higher on the average in every year than their estimates for pupils of other teachers.

It is worth considering the possibility that Miss A's reputation might have had some impact on the marks given by the second-grade teachers, insofar as they would tend to be aware of which children had been with Miss A, and which had come from other classes. If, for example, second-grade teachers were very impressed with Miss A, they may well have expected her children to show more leadership and initiative, and they may have selectively perceived these qualities. By the time the children had reached the third grade, however, teachers were not so likely to have known which child had begun with Miss A and which had been assigned to other teachers. Nevertheless, Table 13 suggests that third-grade teachers saw substantial differences between Miss A's former pupils and those of other teachers—even greater differences than those noted by second-grade teachers. We believe these data support the idea that the development of a positive self-concept in the early school years will tend to perpetuate itself in subsequent levels of achievement, leadership, and initiative.[7]

The permanent record cards provided other opportunities to look at self-concept indicators, but this information is a little more difficult to interpret. Although it is clear that leadership, initiative, and effort are relatively positive, teachers' evaluations of reliability and cooperation could have indicated submissiveness,

[7] We have not discussed the statistical significance of any of the findings in Tables 11, 12, or 13. According to the binomial theorem (Siegel, 1956) the chance of seven plus-minus events having the same direction seven times in a row is .008. The effect observed repeatedly seems a real one.

or compliance; thus, these factors could be indicative of negative as well as positive self-concept.

Table 14 presents specific data on reliability and cooperation. Unlike the previous four variables, reliability and cooperation did not show consistent patterns over a child's elementary-school career. Miss A's pupils appeared to have been high on this characteristic in her own estimation, but the difference virtually disappeared in the last three years of elementary school. Miss A's pupils may still have worked harder, taken more initiative, and achieved higher grades, but they were no more likely to have been "good little boys and girls" than children who were originally assigned to other first-grade classes. Perhaps more than children who are self-assured, children who have little confidence in their abilities may cooperate with their teachers under any circumstance. But, because the pattern of data in Table 14 is not statistically significant, these inferences are completely speculative.

A perennial problem in dealing with data resulting from teacher estimates is the possibility of a "halo effect," such that once teachers decide who are the "good pupils," the tendency may simply be to give them high marks in all personal characteristics. If this is indeed what occurred, then the meaning of the relationships reported in this section may be equally hard to interpret. However, our confidence in the quality of the data being analyzed is increased by the fact that personal characteristics that seem less clearly related to positive self-concept were also less likely to follow exposure to Miss A than variables that do seem closely related to positive self-concept. The different results in Table 14 as compared to those in the preceding three tables suggest that teachers were honestly attempting to make careful discriminations and were not simply entering a string of identical grades for all personal characteristics on each child's permanent record card.

Beyond Elementary School

Although we had many data for our subjects between the ages of approximately five to thirteen, an interesting question is what happened between thirteen and the time we interviewed them in their early thirties. It would have been particu-

TABLE 14
Mean Reliability and Cooperation in Each Year of Elementary School, by First-Grade Teacher

	First-Grade Teacher	
Mean Reliability and Cooperation	*Miss A*	*Others*
Grade I	6.8	5.2
Grade II	6.9	6.0
Grade III	7.2	5.9
Grade IV	6.6	6.3
Grade V	7.3	7.5
Grade VI	6.9	7.0
Grade VII	7.0	6.9

larly useful to know how far these students had pursued their formal schooling, at what age they had left school, and what grade they had completed. Questions during the interviews elicited these facts. An analysis of this information shows that 57 percent of Miss A's children had completed tenth grade or higher (as compared to 42 percent for the pupils of other teachers). This result is intriguing in view of the evidence that length of schooling correlates highly with adult status ($r = .50$). The trends are in a direction that supports our general hypothesis, but the relationship reported here is not statistically significant.

In retrospect we have learned that our interview schedule was especially weak in certain respects: merely asking for length of schooling was not as useful as we had thought originally, because school performance depended on which stream—college preparatory or vocational—pupils were tracked in, as well as the length of time they persisted. Because of this flaw, we are unable to draw conclusions concerning the effect of Miss A on length of schooling as a means of influencing adult status.

The second phase of this study thus demonstrated a dramatic relationship between first-grade teachers and adult status. The analysis of the data supported the hypothesis that a good first-grade teacher can provide children such a major head start, that the effects, in terms of both academic self-concept and achievement, will continue to be felt in later life. Although an attempt was made to analyze the possibility that this relationship was spurious, the observed relationships persisted.

The Third and Final Stage of the Analysis

Up to this point, the pattern of data is still somewhat fragmented. Although the findings from the prior analyses supported our general hypothesis concerning the teacher as significant other, the separate parts have yet to be brought together into a single, comprehensive analysis. It now seems appropriate to offer one coherent model. This, then, is the aim of our third stage of analysis, which attempts to increase confidence in our explanation of the impact of Miss A on adult status.

We have chosen to construct a comprehensive model of these longitudinal processes by using the principles and techniques of path analysis (Duncan, 1966; Heise, 1969). We take this step with the full knowledge of the advantages and disadvantages that accompany this technique. The first major disadvantage is that readers unfamiliar with the technique sometimes incorrectly attribute a quasi-mystical causality to the observed relationships. One should be aware, therefore, that we do not consider the estimation of a path coefficient as a causal statement in and of itself; in fact, the estimation procedure is based on multiple regression and is no more or less valid than any other regression analysis. Causality is a complex methodological and philosophical topic that is certainly beyond the scope of this report. Even in the physical sciences where highly controlled experimentation is possible, the assessment of causation is problematic, and yet it is clearly easier there than in social-science research. Nonetheless, the longitudinal character of our data gives us a definite advantage in making causal statements over cross-sectional research done in the social sciences, where causal attributions

are also sought. In evaluating the results of our research, we urge readers to keep in view all those characteristics habitually used in determining causality: the size of the sample, the generalizability of the research setting, the possibility of correlated but unmeasured antecedent variables, the possibility of reciprocal causation, the methodological biases producing artifactually high correlations, and so on.

Another problem with path analysis is the number of statistical and methodologic assumptions required. These assumptions include all of those used for multiple regression and the general linear model; a high level of reliability of measurement; "undebatable" causal order; and the inclusion of all important variables in the system. The assumptions required for multiple regression are typically robust, and we have not violated them severely. The fact that some of our variables are dichotomous or ordinal in nature is not problematic (Boyle, 1970; Cohen, 1968; Labovitz, 1970). All of the variables have been measured with an acceptable degree of reliability, and we never depart from our undebatable causal ordering. Although it is difficult to prove that we have included all important variables in the system, it is clear, nonetheless, that we have included the major variables, in a fashion comparable to other path analyses.

With these caveats and explanations behind us, we can now exploit the several advantages of path analysis. To begin, we may use this technique to link together the previously disparate analyses into one longitudinal model of a process of intergenerational mobility—an extension suggested by the well-known Blau-Duncan (1967) model. Other extensions of this model of intergenerational mobility have included the examination of such social-psychological variables as ambition and mental ability, as well as the effects of high school and college. None of these extensions, however, has focused on the elementary school. Thus, path analysis achieves our primary goal of building a single model spanning the generation upon which our research was conducted. The ability of path analysis to link longitudinally an entire process is another asset of the procedure. Finally, it allows one to multiply adjacent paths and calculate mediated effects, in order to isolate the mechanism through which Miss A's influence was felt. In addition, the size of all the standardized partial regression coefficients in different parts of the model are directly comparable. This is critically important in evaluating the magnitude of Miss A's effects.

Before continuing, specific differences between this research and other models of intergenerational mobility must be noted. Among other things, the adult-status score reported here is not comparable to other occupational-prestige scores. It is a reliable measure of status, but it is more global (including personal appearance, for example) and more sensitive to social influence such as might be exercised by a first-grade teacher. As a consequence, the variance and sensitivity of adult status are maximized within a narrow socioeconomic range. In constructing the variable "Miss A," we have deliberately contrasted a superior teacher with ordinary teachers. The research setting, therefore, is by no means typical. It should be readily apparent that our data cover only a narrow range of socioeconomic status and make use of a maximally powerful teacher variable and a very sensitive dependent variable.

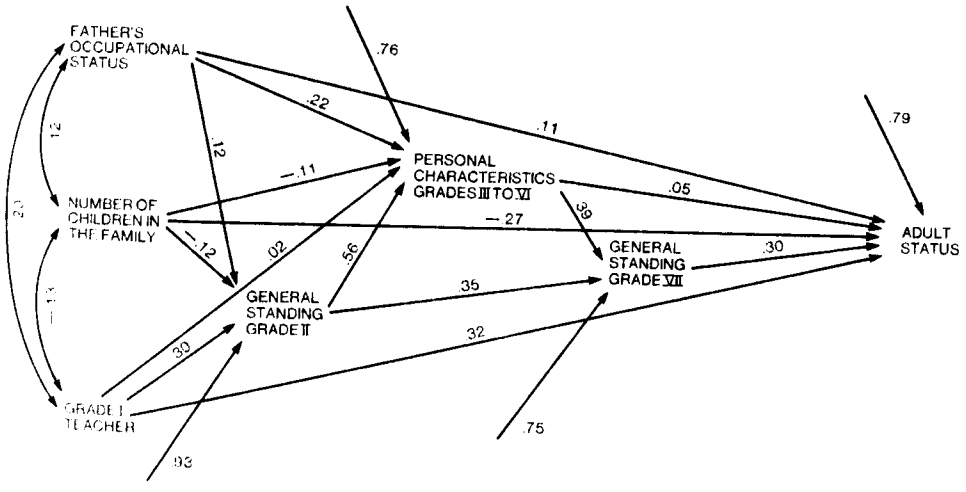
In constructing our path analysis, we have chosen the model that most succinctly expresses in logical fashion the complex relationships involved. Although our model does not represent the most powerful explanation for intergenerational mobility, it does exploit the data for all their informative value and minimizes any weaknesses in them.

The three exogenous variables shown in Figure 1 are identified as the father's occupational status, the number of children in the family, and the first-grade teacher. Of these three variables, the first two were chosen because both have been shown to influence subsequent achievement more strongly than any other background variables, and, therefore, as control variables, they provide the most conservative test of Miss A's influence on adult status. As with previous analyses, the variable "first-grade teacher" was coded in dichotomous form, with respondents in Miss A's class receiving a score of 2 and others a score of 1.

In the center of the diagram, the elementary-school process is represented by two kinds of endogenous variables—general standing and personal characteristics. These variables represent a selection from the complex year-by-year data available to us. Each is a summary index made up from a set of related variables that correlated highly within each set but not so highly across individual sets. General standing is an aggregate measure of achievement in all the individual academic school subjects averaged over an entire year; second grade was chosen because it is the earliest measure independent of the effect of Miss A; seventh grade was chosen because it is the final stage of the elementary-school process, most distant from the effect of Miss A, and closest to adult status.

The personal-characteristics variable is formed from marks for effort, coopera-

FIGURE 1
Path Analysis Representing Dependence of Adult Status on Elementary-School and Background Factors



tion, reliability, leadership, and initiative summed over the years from the third to the sixth grade. Choosing these years between second and seventh grade allowed us to study the interactive relationship between general standing and personal characteristics. The validity of the personal-characteristics variable was strengthened by summing the marks in these areas over four years. In contrast to the general-standing grades, which depend on relatively objective test results, the personal-characteristics variables were simply the teachers' estimates, and these could vary markedly from year to year due to differences in a teacher's perception of a given child or of children in general. Hence, personal characteristics correlate highly within any particular year but not as highly from one year to another, the range spanning from .91 to .48. In contrast, the measures of academic achievement were more stable from year to year, ranging from .62 to .70. Summing personal-characteristics grades over four years reduced the distorting effect of any idiosyncratic teacher estimate; such a precaution, however, was not necessary in analyzing academic achievement.

Adult status, the main dependent variable in this study, is a summative index of six variables described earlier: type of house, rent paid, educational attainment, occupational prestige, appearance of house, and personal appearance of the respondent. The individual correlations between each of these six component variables and adult status range from .50 to .80. An attempt has been made to obtain these variables from independent sources, so as not to create any artificial intercorrelations among variables resulting from same-observer bias. As a consequence, data on a father's occupational status, the number of children in the family, and the first-grade teacher are objective facts obtained from the permanent record cards and corroborated by subjects during the interviews. General standing in second grade has been measured by the reports of achievement from second-grade teachers; personal characteristics from third to sixth grade were estimated by four different teachers for each child; and, finally, general standing in seventh grade was measured by the seventh-grade teacher. Thus, because children were sent to different teachers in each subsequent year, we gained the benefits not only of longitudinal data but also of data from independent sources as well. The causal ordering of variables in Figure 1 is based on strict temporal sequence of occurrence.

The estimation of the model is based on standard techniques of path analysis (Duncan, 1966). Unanalyzed relationships between exogenous variables are simple correlations. The path coefficients in Figure 1 are standardized partial regression coefficients and represent a measure of effect in standard-deviation units. As may be seen, having Miss A as a first-grade teacher noticeably improves academic performance in the next year. The effect, .30, is statistically significant ($p < .05$) and stronger than the effect either of father's occupational status (.12) or of the number of children in the family (−.12). Miss A apparently also has influenced personal characteristics, and Figure 1 suggests that this took place very largely through her impact on subsequent academic performance.

Academic performance in the second grade further appears to have had a powerful influence on personal characteristics in the third to the sixth grade (.56).

The indirect effect of Miss A is .17 (.30 × .56); the direct effect of Miss A on personal characteristics in these years, however, appears to be very small (.02). Academic standing in the seventh grade is the last and strongest measure of elementary-school attainment. This score is influenced not only by academic grades obtained five years earlier (.35) but by personal characteristics in the previous four years (.39). Miss A's influence is constrained to act through these two school variables, but nevertheless, it is still fairly potent: through second-grade performance her effect is .10 (.30 × .35); through personal characteristics her effect is .06 (.30 × .56 × .39); and the total effect of Miss A on seventh-grade academic performance is .16.

All of these variables contribute to adult status. Background factors appear to have moderate effects in the predicted direction: father's occupational status has an effect of .11 and size of family a negative effect of −.27. The personal-characteristics variable seems to have a negligible effect, .05, on adult status. Academic standing has a moderate effect of .30, and Miss A has a direct effect of .32. Overall, the direct effect of Miss A is the strongest influence on adult status, and it is statistically significant at the .01 level. Thirty-eight percent of the variance in adult status is explained by this model.

Summary and Conclusions

This study has focused on the relationship between school experiences and success in adult life. The findings display a positive correlation between one first-grade teacher and the adult success of children from a disadvantaged urban neighborhood. Most importantly, the findings suggest that an effective first-grade teacher can influence social mobility. We have described the process through which a good teacher shapes both the academic self-concept and achievement of the pupil so that an initial foundation yields cumulative benefits in later stages of life. The analysis was strengthened by the use of longitudinal data and independent measurements and estimates of the characteristics and achievements of one group of children.

The findings of this research are at variance with the conclusions of Coleman, Campbell, Hobson, McPartland, Mood, Weinfeld and York (1966) and Jencks, Smith, Acland, Bane, Cohen, Gintis, Heyns, and Michelson (1972), who are pessimistic about the effects of schooling on later adult status. As far as we can ascertain, Jencks and his colleagues dealt largely with this material cross-sectionally—incorporating several sources and aggregate data on different individuals. Although they have not analyzed teacher influence on the later achievement of pupils, they state:

> There is no evidence that building a school playground, for example, will affect the students' chances of learning to read, getting into college, or making $50,000 a year when they are 50. Building a playground may, however, have a considerable effect on the students' chances of having a good time during recess when they are 8. The same thing is probably also true of small classes, competent teachers, and a

dozen other things that distinguish adequately from inadequately financed schools. (p. 29)

These findings have been criticized by Luecke and McGinn (1975), who claim that the application of regression-analysis techniques to cross-sectional data produces misleading results. By drawing attention to their oversights, we are not attempting to belittle the contributions of these researchers but, rather, we are making a plea that an important element of social life such as education should be studied with a broad range of hypotheses and methodologies.

We share many points of agreement with Coleman and Jencks. Like them, we recognize, and our path analysis reveals, the importance of family characteristics for adult status. Our differences with them arise from two factors: first, the particular nature of our data, which focuses on relationships not available in other analyses of the impact of schooling; second, our subjects are drawn from a very narrow socioeconomic range, and thus the relative impact of the teacher is more obvious than it would have been in a study with a sample from a broader socioeconomic range. We do not dispute the fact that, regardless of teacher quality, children from privileged background are more likely to achieve high adult status than children from disadvantaged backgrounds. But our research does show that the teacher can make a difference, not only to pupils' lives in schools but to their future as well.

Our conclusions must be viewed with caution. We have already explained the shortcomings in the size of our sample and the limits on the generalizability of our research setting. We have presented evidence that further work on the relationship between teacher effect and adult status is necessary. In the meantime, teachers currently at work should not accept too readily the frequent assertion that their efforts make no long-term difference to the future success of their pupils. While we must not return to the unreasonably inflated expectations for public education of the last decade, we must also not dismiss the likelihood that teachers' best efforts may have positive long-range results. To do less may destroy teachers' morale and minimize their importance, as significant others, to their pupils.

References

Blau, P. M., & Duncan, O. D. *The American occupational structure.* New York: Wiley, 1967.

Blishen, B. R. The construction and use of an occupational class scale. *Canadian Journal of Economics and Political Science,* 1958, **24,** 521–531.

Bloom, B. S. *Stability and change in human characteristics.* New York: Wiley, 1964.

Bloom, B. S., Davis, A., & Hess, R. *Compensatory education for cultural deprivation.* Holt, Rinehart & Winston, 1965.

Boyle, R. D. Path analysis and ordinal data. *American Journal of Sociology,* 1970, **75,** 461–480.

Brookover, W. B., & Gottlieb, D. *A sociology of education* (2nd ed.). New York: American Book Company, 1964.

Cohen, J. Multiple regression as a general data-analytic system. *Psychological Bulletin,* 1968, **70,** 426–443.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. *Equality of educational opportunity.* Washington, D.C.: U.S. Government Printing Office, 1966.

Collins, R. Miss Apple Daisy. *Reader's Digest* (Canadian Edition), September 1976, pp. 142–147.

Duncan, O. D. Path analysis: Sociological examples. *American Journal of Sociology,* 1966, **72,** 1–16.

Ferguson, G. A., *Statistical analysis in psychological education.* New York: McGraw-Hill, 1959.

Heise, D. R. Problems in path analysis and causal inference. In E. Borgatta (Ed.), *Sociological methodology.* San Francisco: Jossey-Bass, 1969.

Jencks, C., Smith, M., Acland, H., Bane, M. J., Cohen, D. K., Gintis, H., Heyns, B., & Michelson, S. *Inequality: A reassessment of the effect of family and schooling in America.* New York: Basic Books, 1972.

Justman, J. Stability of academic aptitude and reading test scores of mobile and non-mobile disadvantaged children. In Passow, A. H., Goldberg, M., & Tannenbaum, A. J. (Eds.), *Education of the disadvantaged.* New York: Holt, Rinehart & Winston, 1967.

Labovitz, S. The assignment of numbers to rank order categories. *American Sociological Review,* 1970, **35,** 515–524.

Luecke, D. E., & McGinn, N. F. Regression analyses and education production functions: Can they be trusted? *Harvard Educational Review,* 1975, **45,** 325–350.

Merton, R. K. *Social theory and social structures.* Glencoe, Ill.: The Free Press, 1957.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. *Statistical package for the social sciences* (2nd ed.). New York: McGraw-Hill, 1975.

Rist, R. C. Student social class and teacher expectations: The self-fulfilling prophecy in ghetto education. *Harvard Educational Review,* 1970, **40,** 411–451.

Rosenthal, R., & Jacobson, L. *Pygmalion in the classroom.* New York: Holt, Rinehart & Winston, 1968.

Schwebel, A. I., & Cherlin, D. C. Physical and social distancing in teacher-pupil relationships. *Journal of Educational Psychology,* 1972, **63,** 543–550.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Thorndike, R. L. Review of *Pygmalion in the Classroom* by R. Rosenthal & L. Jacobson. *American Educational Research Journal,* 1968, **5,** 708–711.

Thorndike, R. L. But you have to know how to tell time. *American Educational Research Journal,* 1969, **6,** 692.