DOCUMENT RESUME

ED 102 437 CG 009 427

AUTHOR Entwisle, Doris R.; And Others
TITLE Expectation Theory in the Classroom. Final Report.
INSTITUTION Johns Hopkins Univ., Baltimore, Md. Dept. of Social Relations.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
BUREAU NO 1-0034
PUB DATE Jun 74
GRANT OEG-3-71-0122
NOTE 437p.

EDRS PRICE MF-$0.76 HC-$22.21 PLUS POSTAGE
DESCRIPTORS *Behavior Development; Elementary Education; *Elementary School Students; *Expectation; Grades (Scholastic); *Individual Development; *Performance; Research Projects; Theories

ABSTRACT

The purpose of this research was to study expectations of elementary school children in two ways: experimentally and observationally. Expectations may be roughly defined as a child's ideas of his own ability at a particular task. From the data it appears that childrens' expectations could be raised experimentally by a suitable adult and high expectations in one area generalize into other unrelated areas. The experiments are summarized in a number of published articles reproduced herein and listed in the bibliography. The observational data focus on children in first and second grades in a white middle-class school and in an integrated lower class school. From the time they enter school individual children are followed to see how their expectations for their own performance in reading, arithmetic, and conduct develop. Their expectations, and their parents' expectations for them, are repeatedly measured. Children in both places have, on the average, very high expectations for themselves before they get a report card, higher than their parents. These expectations do not diminish much when marks are lower than expected; in fact for the majority, expectations are maintained over first grade. Children whose marks improve are likely to be those whose expectations exceeded marks. (Author)

Final Report

Project No. 1.0034

Grant No. OEG-3-71-0122

EXPECTATION THEORY IN THE CLASSROOM

Doris R. Entwisle
Murray Webster, Jr.
Leslie A. Hayduk

The Johns Hopkins University
Baltimore, Maryland 21218

June, 1974

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE
Office of Education

Abstract*

The purpose of this research was to study expectations of elementary school children in two ways: experimentally (Section I), and observationally (Section II). Expectations may be roughly defined as a child's ideas of his own ability at a particular task.

The experiments focus on (1) how children's expectations for their own performance may be raised, (2) what kinds of persons can raise children's expectations, (3) what kinds of children are susceptible to expectation raising, (4) whether children's expectations generalize from one area to another, (5) whether children express behaviorally their expectations for other children, and (6) what kinds of data shape children's expectations for other children. Children's expectations can be raised experimentally by a suitable adult and high expectations in one area generalize into other unrelated areas. The experiments are summarized in a number of published articles reproduced herein and listed in the bibliography.

The observational data focus on children in first and second grades in a white middle-class school and in an integrated lower class school. From the time they enter school individual children are followed to see how their expectations for their own performance in reading, arithmetic, and conduct develop. Their expectations, and their parents' expectations for them are repeatedly measured. Children in both places have, on the average, very high expectations for themselves before they get a report card, higher than their parents. These expectations do not diminish much when marks are lower than expected, in fact for the majority, expectations are maintained over first grade. Children whose marks improve are likely to be those whose expectations exceeded marks.

---

* The reader should also see the "Overview", p. xiv.

Preface

This research has been conducted over a 3-year period. It had its genesis in work done even earlier, starting with exploratory studies in 1969. There is a clear intellectual debt owed to Joseph Berger and his associates at Stanford University who have formulated and tested Expectations States Theory in the laboratory. In a very real way our research represents the extension of that research into field settings. The theory led us to work directly with children's expectations. It also led us to some observational work, to see what events in young children's academic careers lead them to formulate high (or low) expectations for themselves.

Like most research, that reported here probably raises more questions than it answers. That it can answer any is due in a very real sense to the cooperation and patience of teachers and children in schools in both Baltimore County and Baltimore City. Where possible, individual principals and others are identified in reports reproduced for inclusion in this Final Report. In the case of Section II, which covers the observational work, confidentiality requires that we do not identify the particular schools involved in the study. We acknowledge here the splendid cooperation we have received from these schools, and regret that they must remain anonymous even here.

A number of graduate students have aided in various phases of the work: Margaret Boeckmann, Ellen Dickstein, Esther Greif, Guillermina Jasso, Susan Doering, Muriel Berkeley, Marguerite Bryan. Other persons who have aided in the procuring and analysis of data are Barbara Bricks, Laura Gordon, and Judy Kennedy. Eileen Rudert is also due thanks.

A very special acknowledgment is due to Linda Olson who for the past two years has assumed a very heavy role in this project, and contributed to every part of it.

Doris R. Entwisle
Murray Webster, Jr.
Leslie Hayduk

Baltimore, Maryland
June, 1974

TABLE OF CONTENTS

Pages

LIST OF TABLES

Section II

Chapter 3

Table | Page

Table | Page

Chapter 4

Chapter 5

Table

Table

Table | Page

Table | Page

Table | Page

LIST OF FIGURES

OVERVIEW

This final report is divided into two sections, Section I: Experimental Studies of Expectations and Section II: Observational Studies of Expectations. Each section deals with a series of studies on children's expectations.

Section I is devoted to experimental studies and consists mainly of a series of papers which present and summarize research supported by this grant. These will be numbered serially to facilitate classifying. Three papers have already appeared in print, (1) "Raising Children's Performance Expectations," Social Science Research, 1972, 1 (June): 147-158, (2) "Research Notes: Status Factors in Expectation Raising," Sociology of Education, 1973, 46 (Winter): 115-126, (3) "Raising Children's Expectations for Their Own Performance: A Classroom Application," Chapter 7 in Expectation States Theory, J. Berger, T. Conner, & M. H. Fisek (Eds.), Cambridge: Winthrop Publishers, Inc., 1974. (4) "Expectations in Mixed Racial Groups" will appear in the summer in Sociology of Education, and (5) "Expectation Effects on Performance Evaluation" has been submitted for publication. Others ("Teacher Expectancies and Student Expectation States," Pacific Sociological Association, April 1972; "Raising Expectations in the Classroom," Eastern Sociological Association, April 1972; "Status Factors in Expectation Raising," American Sociological Association, August 1972; "Expectations in Mixed Racial Groups," American Educational Research Association, February 1973; "Development of Expectations in Mixed Racial Groups," American Psychological Association, September 1974) have been presented at professional meetings and overlap to some degree the papers in print.

These papers have appeared throughout the term of this research project, and each summarized a particular series of field experiments with grade school children. The experiments focus on (1) how children's expectations for their own performance may be raised, (2) what kinds of persons can raise children's expectations, (3) what kinds of children are susceptible to expectation raising, (4) whether children's expectations generalize from one area to another, (5) whether children express behaviorally their expectations for other children, and (6) what kinds of data shape children's expectations for other children. All of the experiments are concerned directly with children's expectations.

The first stage of the experimental research was to analyze classroom problem solving interaction from a theoretical perspective which defined the relevant variables and predicted their interrelations. Experiments then were conducted to demonstrate the utility of our theoretical analysis; to show, that is, that we could predict the effects of specific types of experimental intervention in naturalistic classroom studies. Each experiment focuses wither on a significant link in the theoretical process, or upon elaborating subpopulation differences revealed in a previous experiment.

Section II is devoted to observational studies and presents longitudinal data on expectations of children in three cohorts: (1) white middle-class children who entered first grade in 1971; (2) white middle-class children who entered first grade in 1972; (3) white and black lower-class children who entered first grade in 1972. From the time they entered school these children have been followed to see how their expectations for their own performance develop and change. The children's expectations for their marks in three subject areas (Reading, Arithmetic, and Conduct) have been tabulated twice each academic year (before the first report card in the fall term and just before the end of the spring term). In addition other data characterizing the children (IQ's, marks in Reading, Arithmetic, Conduct, self-esteem test scores, and the like) have been studies over time to see how children's expectations relate to them (e.g. in the case of IQ), or how children's expectations respond to them (e.g. in the case of marks).

The child has been studied within his social network. Parents' expectations, some teacher forecasts for achievement, and the child's sociometric standing in his own class have been studied in relation to the child's expectation level. Besides portraying individual children and cohorts over time, the observational data can be aggregated by cohort, or in the case of the two middle-class cohorts, across cohorts as well for the first grade year. (Further work is planned continuing with these cohorts until the end of third grade.) The aggregation permits examination of a number of important problems, for example, whether male and female children are the targets of differential expectations by parents for achievement in reading or arithmetic. Or, whether black children and white children have different expectation levels or have different patterns of expectation change over time.

The work is novel in several respects. Expectations of children in the first four grades of school have not previously been studied. Cohen and her associates have carried out some experiments with junior high students (especially blacks), and Brookover and his associates have carried out a longitudinal study of students' "self-conception of ability" lasting from the 7th to the 12th grade which included some experimental phases. But other than these two efforts, there are no studies aimed at goals similar to the goals of ours, and both of these cited studies concern much older children. One need not labor here the difficulties involved in working with children below the age when they can respond easily in writing using group procedures, or when they are easily interviewed. But a major novelty in our work springs from the tender age of the subjects.

Another novel feature is that the actual process of expectation formation has not been studied previously. Expectations for school performance, or related variables like aspiration level and locus of control, are often studied in junior high or high school age subjects but no one has ever inquired as to how these expectations are initially established. In previous research expectations are assessed which already exist at the time a study begins. This is important for two

reasons. First, there is considerable sentiment that by third grade (the terminal level for most of our work) children's performance levels are established and will not change much for the remainder of their school career (Kraus, 1973; Husen, 1969; Bloom, 1964). Thus instead of viewing expectations as a characteristic which has potential for change, expectations are viewed as more or less fixed, like IQ. Second, expectations in their early stages of formation lead to feedback loops--expectations determine performance which then determine expectations. These loops may become so functionally autonomous as to defy intervention in later childhood.

Still another novel aspect is the comparison by social class level of the effectiveness of various kinds of agents as expectation raisers and the impact of feedback from different persons on expectation level. To our knowledge no one before has looked into how racial mismatch between teachers and primary age children may affect performance, and there are few data on how social class mismatch between teachers and children affects feedback on performance. In one set of experiments (4) we find that black adults are relatively ineffective at raising expectations of white lower-class children although they are effective with black lower-class or white middle-class children. We also find in the observational study that the range of marks used by first-grade teachers is very different between the middle-class and the lower-class schools. Feedback in the way of report cards implies rather different things to children in these two locales.

A still further source of novelty is that, with the exception of the Brookover and Cohen projects refered to above, work directly with children's expectations, rather than indirectly with teachers' expectations, has not been the topic of research in "expectancy" studies. A shower of recent studies concern teacher expectancy or Pygmalion effects. These have two aims: (1) demonstrating that expectations teachers hold for pupils shape pupil performance, or (2) pinpointing teacher behaviors that mediate the expectancy effects.

A number of studies give teachers false information that some children are "gifted" or "late bloomers" when in fact the students so designated are randomly selected. Then the children's subsequent performance is evaluated. A review of the statistical evidence on the Pygmalion effect, given in Rosenthal and Rubin (1971), suggests that about 39% of such maneuvers are successful. To explain unsuccessful maneuvers one can surmise that in some cases teachers pay no heed to the "information" given them or may forget it. Also apparently teachers act out their expectations in different ways, determined at least in part by the characteristics of the students with whom they are interacting. Our experimental work goes far toward elucidating how the actual process of how teacher expectations affect performance. This is discussed in detail in both (3) and (4).

Studies that seek to understand exactly how teachers behave to cause "high-expectancy" pupils to perform better are less numerous but more important in terms of implications for practice. The Brophy and Good (1970)

research, for example, reveals that teachers demand more from students for whom they hold high expectations. Closely related are "ex post facto" naturalistic studies like the recent one (Seaver, 1973) showing that younger siblings are rewarded to some extent on the basis of teachers' prior experience with their older brothers and sisters, or the Palardy (1969) study showing that those teachers who expect boys to be slower readers than girls have such boys by the end of first grade.

Anyone would suppose, however, to operate on students' expectations directly is the most effective way to mobilize expectations toward educational goals. A severe limitation to the Pygmalion work, if its implications are to be followed in applied work, is that deception is involved. To tell teachers that some students are "gifted" may be defensible to demonstrate an effect, as a tour de force, if you will. But to give teachers false or misleading information about pupils is hardly a practice that can be countenanced in everyday affairs. To tell teachers that some students are "gifted" furthermore implies "more gifted than others," or that others are not gifted. This negative implication we have never seen explicitly mentioned, although one must presume, since attention and reinforcement are distributed, that the pupil's falsely designated as non-gifted are perhaps denied the attention they need to flourish. The important point, however, is that presumably pupils' expectations for themselves are shaped by teachers' expectations. That being the case, pupils' expectations may be a more potent and therefore relevant variable to serve as a research target.

In an important way our research fuses or expands upon a number of important methodological approaches. The experimental portion of the research took its impetus from tight, carefully controlled, laboratory studies where the dependent variable was susceptibility to influence in cases of disagreement. Will an individual, for example, change his response to an ambiguous stimulis when confronted with a disagreeing alter for whom he previously holds high expectations? Expectations in laboratory work have uniformly been the topic of the experimental maneuver rather than the dependent variable. Expectations are manipulated. Subjects are told fictitiously high scores or given false information about alter's status so they will hold high expectations for him compared to themselves.

It is quite another matter to cause expectations to change by some sequence of believable events rather than by fictitious assignment. Also it is quite another matter to look for a behavior which the subject generates spontaneously, like volunteering to perform, rather than to ally responses to an experimenter-imposed task like judging slides. Resisting or accepting influence which is artifically generated is much simpler to arrange, and much less like an ordinary life circumstance, than is deciding on one's own initiative to perform or to volunteer. We do not wish at all to denigrate the laboratory work but we do wish to point out that laboratory research on expectation problems, although valuable, has constraints that limit its value for predicting or controlling behavior in true-to-life settings. Expectations for the self

or for others arise gradually (we think) in social commerce. They then affect performance and evaluations, both key variables for future expectations and performance. While there may be instances when resolution of disagreement is involved, as say when a teacher and a student interpret a symbol in a poem differently and the student subsequently accepts the teacher's interpretation, this consequence of expectations is a relatively minor one from our viewpoint. The consequences of expectations in terms of activity closely related to learning like performing or volunteering to perform are major consequences in our view and the ones we have studied.

In another way the experiments present a novel methodology. All subjects are maintained in a group context. Expectations by their nature are partly distributive, awarded by persons as groups undertake important tasks. Thus it makes no sense to speak of "high expectations" in any absolute sense. One can speak only of "expectations high with respect to this group" or "expectations low with respect to this set of competitors." Failure to preserve social context has led to considerable confusion about the self-esteem and self-image of blacks or other disadvantaged groups. Study after study has appeared demonstrating that minority group children usually have as high, or higher self-esteem, than majority children. This result has been perplexing because it was thought that the self-view would reflect the society's view of the minority group the self belonged to. Work generally on children's self-esteem or self-concept has pretended the individual exists in social isolation. The present work treats children nested in social groups. Volunteering in our experiments is measured with respect to others in the same group. Likewise in the observational work self-esteem or expectations is assessed in terms of the child's position in his classroom or grade. The importance of the group context cannot be stressed too much. In Cohen's experiments, for example, expectations of blacks for their own performance or of whites for blacks' performance are low in mixed racial work groups. In these experiments the blacks are brought from all-black settings. Our experiments with younger children in an integrated school occur in a mixed racial environment so no change in context occurs. Under these conditions blacks and whites both appear to have high expectations for blacks. Different contexts and contextual effects cannot be ignored.

Perhaps the most significant methodological contribution of this research is in the observational portion. It combines the life-span developmental approach typical of longitudinal studies in psychology with the causally complex structural equation approach characteristic of so much recent sociological research, especially research in occupational or status attainment.

Many psychological studies over time have the intent of elucidating progression of growth, the exact nature of growth curves and the like. Oden (1968), for example, recently brought the Terman gifted sample up to date. Gifted children were traced through adulthood, their

physical and mental health and a host of other variables were observed. One cannot say, however, what factors lead to the superior mental health of the gifted. Certain dependent variables are repeatedly measured in the same individuals over time, and the status of the target group on these variables is compared with that of the general population. But very little in the way of exact causation can be firmly stated from this typical example of longitudinal studies in psychology.

In recent sociological studies over time, by contrast, one or a small number of dependent variables is measured at one point in time, with a few independent variables measured at a single prior point. For example, Alexander's work (1973), using data gathered earlier by Eckland, examines occupational attainment in 1972 of a sample of 1-2000 persons who were interviewed in 1955 when they were either sophomores or seniors in high school. The 1955 interviews assess peer plans, parents' aspirations, family background, and father's occupational status. The object is to account for occupational attainment of the aggregate sample in 1970 on the basis of the set of variables evaluated in 1955. There is no interest in predicting that a certain individual will follow a particular life course or attain a particular status. Rather mobility of the group is measured as a function of the prior variables, and causal chains are sought.

As mentioned, the present observational study draws from both these traditions and is, so far as we know, the first attempt to combine the two for a single data set. Individual children are repeatedly measured over time to see what course their self-expectations take (like the life-span psychological studies). But measures of parent and peer expectations are taken to see what effects appear later in the child's expectations that can be causually related to variables measured earlier (like the causally complex sociological studies). Expectations of black children can be measured, for example, in order to study the impact of earlier feedback from teachers and parents and to assign weights to the importance of feedback from various sources. These matters will be extensively discussed later in Section II.

To sum up: Two sorts of work will be presented, one experimental and one observational. They are conceptually closely related. The experimental studies are carefully controlled attempts to raise children's expectations under different conditions. The observational studies are focused on what events occur naturally to raise (or lower) children's expectations. All the work concerns children in the earliest school grades. Although expectations are presumably most malleable at this time, no other research has previously been concerned with this variable at this age.

SECTION I

EXPERIMENTAL STUDIES OF EXPECTATIONS

Doris R. Entwisle and Murray Webster, Jr.

Introduction

The purpose of this section is to describe a sequential series of experimental studies of the determinants and consequences of performance expectations in classrooms. In line with U.S.O.E.'s goals "to produce and disseminate knowledge," we have published, or are preparing to publish, reports of all of these studies. Consequently this narrative omits many technical details, theoretical qualifications, and unusual or atypical cases in the interest of presenting a coherent view of the research program. We have provided references to the more detailed papers, which are appended to this report.

By the time this grant became active in June 1971, we had already investigated in a preliminary way some of the theoretical issues described here, and had conducted some pilot studies of the basic experimental design. Although this early work was exploratory and many refinements in our design and our thinking were necessary before the program produced useful research data, the contributions at the early stages were vital. NIMH grant No. MH-18183 and The Johns Hopkins University Center for the Social Organization of Schools provided funding of the early stages.

At every step of the way our investigation is closely linked to an explicit theory. By "explicit theory" we mean a set of propositions explicitly stated, with a concern for parsimonious, consistent use of concepts and logical consistency of the propositions; and some independent empirical tests of derivations from those propositions. This feature is an unusual one in educational research, and we see it as a major strength of our program. Among the advantages of this "theoretically informed strategy" are the following. (1) The theory isolates crucial variables from incidental features of a situation; for example, it tells us to focus upon children's expectations for themselves rather than teachers' expectations for children. (2) The theory gives a way to analyze previous findings and points out relations between findings from very different concrete settings; for example, it shows similarities between the "Pygmalion" studies, our classroom experiments, and a large number of small group laboratory studies. (3) The theory explains some findings and predicts others; for example, in some of our publications we analyze the "Pygmalion effect" and show how both successful and unsuccessful attempts to produce it may be explained--in contrast to the mysterious nature this effect assumes when it is mentioned by other researchers. (4) The theory guides research by predicting effects of contemplated intervention techniques; thus all experiments reported in this section use direct operationalizations of the theory.

The theory which guides these experiments is presented in Webster and Sobieszek (1974a, 1974b) and in Entwisle and Webster (1974a).

Basically, it is concerned with the relations of evaluations of performance and ability conceptions. The theory is sketched in the paragraphs which follow.

In the experimental studies our concern was to develop substantively important applications of previously confirmed parts of the theory, not to test the theory itself. The strategy of applying only what we considered reliable derivations enables us to concentrate upon developing practical techniques and measures, not upon theoretical development and reformulation. In other words, because we used only independently confirmed theory to guide our research, we could concentrate attention upon measures, experimental design, and intervention techniques when we got unexpected results; not on the propositions which led to the disconfirmed prediction. In a sequential research program, just knowing where the problem lies when things go wrong is exceedingly valuable.

Expectations in the Classroom

An elementary school classroom may be viewed as a social situation in which evaluations of performance, ability of actors' conceptions of their own and each other's abilities, and chances to perform are of central interest. Thus the primary relevant components of normal problem solving interaction are the following: action opportunities or socially distributed chances to perform, such as those the teacher distributes to children by calling on them; performance outputs or problem solving attempts, such as children make when they offer an answer; positive or negative unit evaluations of individual performance outputs, which are communicated by the teacher and also by other students. All of these, as well as other components not mentioned here in the interest of brevity, are distributed among actors as a direct function of the expectation state associated with each actor. The higher the expectations held for a given actor, the more likely he is to receive an action opportunity, to accept a given action opportunity and make a performance output. Thus, although expectation states are themselves unobservable, they are theorectically assumed to affect the social distribution of all relevant components of task focussed interaction.

Expectation states are persistent--though not unchangeable--conceptions of ability which arise as a result of social interaction. It is crucial to note that expectations are always relative: an individual holds expectations for himself relative to particular other individuals and at a particular task. (This is one reason it makes no sense to speak of the "low self-esteem of blacks" without specifying comparison others and a particular task.) In fact, it is partially due to this relative conceptualization that we are able to raise expectations of children in various social situations using a fairly simple intervention technique.

Expectation states are formed in either of two basic ways, though there are variations on each. First, a series of unit evaluations of a particular actor's performances will at some point crystallize into an expectation state held for him. We assume, in other words, that after

a series of positive unit evaluations a child will move from saying to himself "I think I was right" to "I think I have high ability at this task." At about the same time, of course, other children will make the comparable cognitive change: from "I think he's right" to "I think he has high ability." The only factor that matters for expectation formation is the positive or negative unit evaluations of performance. We call this the "basic evaluation-expectation process."

Second, in the absence of evaluative information relevant to the task, actors may form expectations on the basis of any known evaluated characteristic possessed by the individual. The known evaluated characteristic may be another ability--such as when a teacher says "Bill is pretty good at reading, so he probably can spell well"--or it could be a status characteristic--such as when a child says "Most blacks are dumb." We call this the "spread of relevance process," since a previously non-related characteristic becomes relevant to expectations held for particular actors in this case.

Whichever way expectation states for actors come to be held, they tend to persist unless acted upon. There are also reasons why expectation states tend not to be acted upon or changed. The main reason is that once expectation states for actors are formed, they affect the distribution of unit evaluations, and thus act to reinforce the very conditions which led to their formation. This circular evaluation-expectation-evaluation process is what we call its "self-maintaining feature."

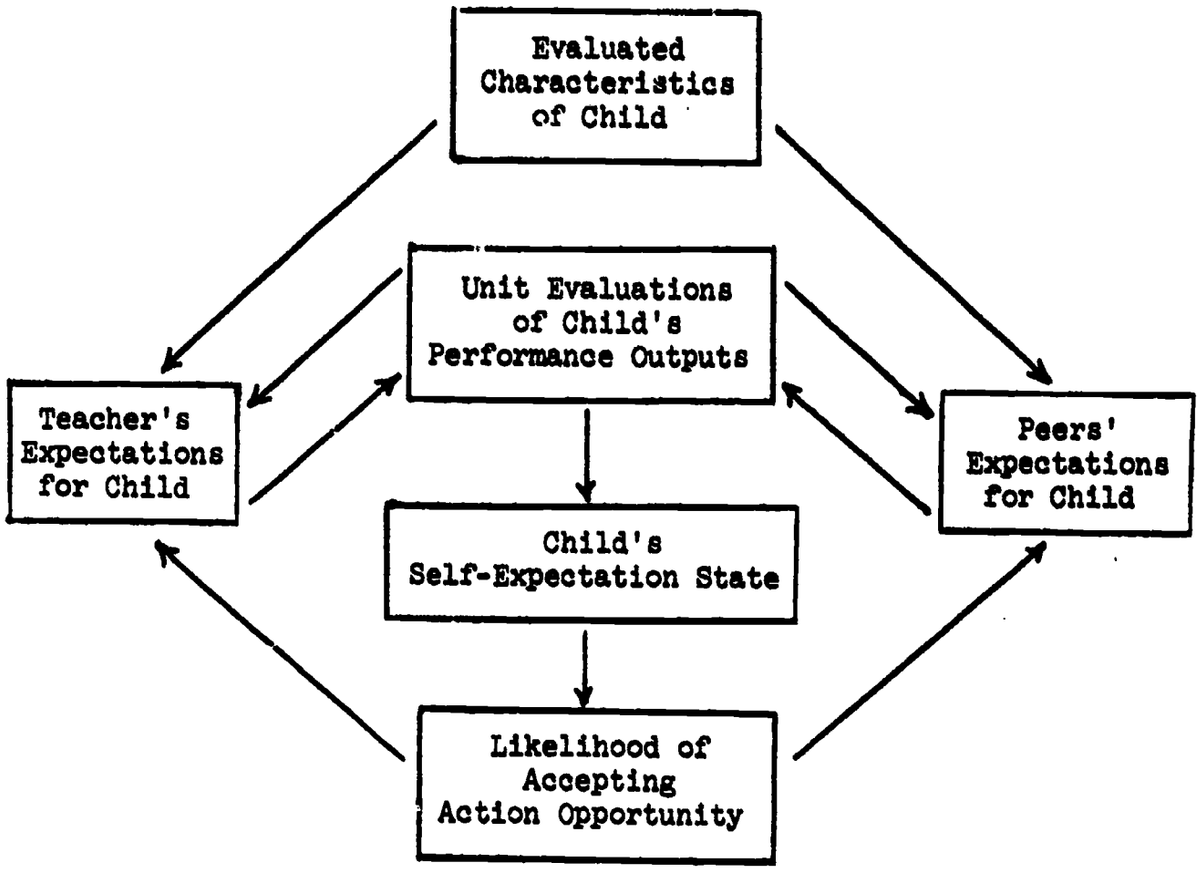
Figure 1 represents the basic expectation-evaluation process, the spread of relevance process, and the self-maintaining feature of expectation states within the classroom. In interpreting the flow chart, assume that influences (represented by arrows) going upwards are weaker in effect than those going downwards. (Figure 1 is intended for illustrative purposes only; in particular, it should not be regarded as an incomplete path analytic diagram.) In Section II of this Final Report we expand the focus to include sources of expectation effects external to the immediate classroom, as well as some attempts to develop quantitative estimates of degrees of effect.

Both for interpreting Figure 1 and in reading the following descriptions of experiments the reader should bear in mind that (1) we are centrally concerned with determinants and consequences of individual children's expectation states, and (2) expectations are always relative to a particular task and to specific comparison others.

First Experimental Series; The Basic Expectation-Raising Experiment

Initial experiments were concerned (1) with demonstrating some crucial steps in our theoretical analysis, particularly the way in which children's expectations are formed and the importance of the child's expectations, and (2) with developing a means to produce results of practical significance. What we settled on to satisfy the second goal was a behavior generally accepted as important in classroom learning, willingness to

Figure 1
Determinants and Consequences of
Classroom Expectation States



engage in new tasks and problem solving attempts, or to answer questions in class. Such behaviors are labelled "acceptance of an action opportunity" in the terminology of expectation states theory, and they should vary directly with the child's (relative) expectations for his ability. Also a part of the second goal was to devise a procedure simple enough to tell someone else--perhaps a teacher who is not particularly interested in experimental design or abstract theory--how to produce the same practical results.

The first goal, confirming the correctness of our analysis, was tackled by inventing an experiment to raise children's expectations for a classroom-like task and setting. High expectations, according to the theory, can be produced by the basic evaluation-expectation process through giving children heavy doses of positive unit evaluations of performances. The combination of goals for these experiments imposed considerable contraints on the experimental task and situation; for example: (1) the task must permit a series of discrete, controllable, plausible unit evaluations; (2) it must resemble classroom activity, but it may not be a task for which children already have expectations; (3) it must not rely on telling either teachers or children something false about their abilities; (4) it must be interesting enough to capture the children's interest, and (5) it must permit relatively objective measurements of the independent and dependent variables of the study.

As the list of criteria suggests, considerable time and effort were required in pilot studies to develop a suitable task and design. The task finally developed is a modification of a story telling task used earlier for research on cognitive development (see Entwisle et al., 1970, appended). The task uses "story skeletons" composed of sentences with missing words, which children supply. A sample story skeleton is shown with supplied words underlined in Figure 2. The experiment has three phases. In Phase I, groups of children--initially 6, but later groups of 4--are assembled and told the experimenters are looking for children who can tell good stories. Stories are to be made up by teams, and the best team stories will be awarded prizes. In each team, the experimenter will read a sentence, and pause when she comes to a blank. Children then raise their hands <u>if they can think of a good word</u>, and one child will be called upon to give the "team's word." The more good words, the better the story, and the better the team's chances to win.

Phase I and Phase III are measurement phases: expectation level is assessed by noting unobtrusively how often each child raises his or her hand. Hand-raising is our measure of relative expectation states for "story-telling ability" in that particular group. In Phase II, the attempt is made to raise one (experimental group) child's expectations by having him make up a story all by himself while the other (control group) children are in another room. Every single performance is given a positive evaluation in Phase II. This procedure, if successful, should raise the child's self-expectations. In Phase III the groups are reconstituted, experimenters are rotated so they do not know who the experimental group child is, and the measurement is repeated. Figure 3 summarizes the basic expectation raising experiment.

Figure 2
Sample of Story-Telling Task

THERE WAS ONCE A ___(very tall prince)___

WHO HAD A ___(castle)___

THAT (HE, SHE) ___(lived in)___

ONE DAY (HE, SHE) HAD TO GO TO ___(the dungeon to see his prisoners)___

(HE, SHE) DID THIS VERY ___(angrily)___

BECAUSE (HE, SHE) WANTED TO ___(make sure they were there)___

THIS WAS VERY DANGEROUS BECAUSE OF THE ___(strong prisoners)___

WHO (WHICH) WAS (WERE) VERY ___(mean)___

IN ORDER TO FOOL THE (FILL IN) THE (FILL IN) DRESSED UP

AS A ___(another prisoner)___

IN SUCH A DISGUISE THE (FILL IN) LOOKED ___(mean)___

AND WHEN THE (FILL IN) SAW THE (FILL IN), THEY ___(welcomed him)___

THIS MADE THE (FILL IN) ___(feel pretty good)___

AND ___/COMPLETE STORY/ (he let his new friends go).___

Figure 3
Summary of the Basic Experiment

	Phase I	Phase II	Phase III
Control Ss	One story is produced (12 words); no evaluations; level of volunteering observed	Control Ss have story read to them by another adult, from 12 to 15 control Ss join in one group	Repeat Phase I, with same control Ss and experimental Ss as in Phase I; experimenters are rotated so the experimenter is unaware of identity of experimental Ss
Experimental Ss	One story is produced (12 words); no evaluations; level of volunteering observed	Experimental Ss make up story individually with the same experimenter they have seen in Phase I; receive all positive evaluations; experimenal Ss join control Ss at end of this phase	Repeat Phase I, with same control Ss and experimental Ss as in Phase I; experimenters are rotated so the experimenter is unaware of identity of experimental Ss

Data are tabulated by calculating gain scores for children between Phase I and Phase III. This procedure allows each child to act as his own control for such extraneous sources of change as excitment, growing familiarity with the situation, etc.

The first experiments were performed with children from two very different settings, two black inner-city schools and two white rural schools. Children were third and fourth graders, both sexes (n = 272). In all groups, the mean gain score of experimental group children exceeded mean gain for control group children. Thus we concluded that we had some success at meeting our initial goals. However, the procedure was less successful with the black children than with the white, and was less successful with fourth graders than with third graders. (For detailed references see Entwisle and Webster, 1972, 1973 (all appended). These reports constitute the exploratory segments of the project.)

Second Experimental Series

The second experimental series was designed to explore the generality across sociological subgroups of the experimental findings that children's expectations could be raised at classroom-like tasks. Data were gathered from children in grades one through four of a white suburban school. Comparison of data from the first and second experimental series permits us to assess the effects of age and residential locus, as well as race. The total number of children in the second series is 381.

Results again indicated success at expectation raising. For all four grades of the suburban school, comparisons with earlier data showed no significant effects by grade or residential locus, and no interactions involving any of these three factors.

These data together with the first experimental series provide strong evidence for the generality of the expectation raising effect, and for wide applicability of the experimental maneuver devised. Although not all possible groups of children have been studied (for example, no black middle class children have so far been included), the sociological range is great enough for us to conclude that the phenomenon is a general one. (See for detailed references Entwisle and Webster, 1973.)

Third Experimental Series

The next set of experiments was directed to some special problems which arose during the course of the earlier experiments. The third series investigated issues of theoretical analysis, of experimental methodology, and of several alternative interpretations of our results.

First, we wished to explore reasons for the relative lack of success at expectation raising with inner-city black children; especially, fourth graders. Although as earlier noted we were able to raise expectations

somewhat in this group, the effect was not large enough to attain statistical significance. We posited two possible reasons for the relative lack of success. (1) Attending a black ghetto school is often seen as causing children to learn that they will not succeed. In our terms, black children receive much negative evaluation, and come to hold low, fixed expectations for themselves. Our relatively simple experimental procedure was then simply not sufficient to convince these children that they had high ability. We dubbed this the "debilitating effect of school" hypothesis. (2) An entirely different line of explanation, suggested by Katz and his associates (Katz, 1970), is that some black children perceive white adults (such as our experimenters) as hostile and insincere; in our terms, perhaps the children did not believe the positive evaluations they were given. One might term this the "race-credibility gap."

To distinguish between these interpretations, we returned to the same inner-city school a year later with black experimenters, and replicated the earlier experiments with black inner-city children. The total number of children in these repeat experiments was 76. Results indicated that black adults were very successful at raising expectations of black inner-city children even though previous attempts with similar children by white experimenters had been relatively unsuccessful. From this we were able to reject the debilitating effect of school hypothesis, and to accept the suggestion contained in Katz's work. We also were sensitized to the significance of racial match or mismatch in these experiments, a question we pursued at length later.

Second, we were interested in the comparative lack of success with white suburban girls. We replicated the experiments with third and fourth grade girls at a second, equivalent suburban school ($n = 60$). Again we were unable to produce statistically significant differences between gain scores for experimental and control groups, although the gain of experimental groups was consistently positive. We tentatively concluded that a "contrast effect" may have produced the failure: girls, especially middle class suburban girls, are rather accustomed to receiving praise and positive evaluation for their work. Our experimental technique contrasted against this uniformly positive background is not so impressive to girls as it is to boys, or to rural or inner city children whose general background is less positive.

Third, in discussing our work with colleagues, an alternative "behavioral modification" interpretation of the effect of the experiments was offered. Essentially such an interpretation argues that children in the experimental group increase the rate of emission of a behavior which is reinforced, rather than raise their expectations. The links to an S-R versus a cognitive theoretical stance are obvious. Although we are not interested in ruling out every possible competing explanation and the ease with which alternative explanations can be invented makes that activity fruitless, we felt the behavioral modification interpretation usually does not include any notion of cognitions of the individuals. By contrast, it is central to the expectation theory interpretation that a cognition--a self-referent idea--gets changed. We see behaviors as the direct consequence of cognitions.

On analytic grounds, we do not find the behavioral modification interpretation persuasive: it requires a long chain of untestable assumptions regarding how variables in our experiment are to be interpreted. The experimental children had additional "practice" telling a Phase II story, and they had more "personal contact" with the experimenter. More telling however is the fact that the actual behavior whose increase is measured--hand-raising-- is never reinforced. We evaluate (or reinforce) words spoken, not handraising activity. Still, some simpler variants of the behavioral interpretation could be offered.

To assess the status of these ideas we conducted some "special control" experiments. In Phase II two children were selected: one to receive the standard expectation raising treatment, and a second (special control) child who made up a story by himself with an experimenter just as the experimental child did, but without any evaluations of his performances. Insofar as possible, the special control child received identical experiences to those received by the experimental child excpet the crucial one, unit evaluations. Results of these experiments (n = 88) show essentially zero changes or both control and special control children, and a strong significant effect for the experimental group children. We conclude these results are interpretable only from our expectation raising perspective. (For detailed references see Entwisle and Webster, 1974a).

Fourth Experimental Series

The purpose of the fourth experimental series was to study in more detail the effects of racial differences in formation and change of expectations. In particular, we were concerned with (1) the effects of racial match or mismatch between experimenters and subjects; and (2) the effects of the basic experiment using integrated groups of childen. All experiments to this point had used same-sex and same-race children in experimental groups. The first issue has implications for deciding who can be an effective source of expectations for children, and also for deployment of teachers in school systems. The second issue is relevant to the effects of school integration upon ability conceptions of children.

(1) Results of experiments with children of the same race in each group show, generally, that white experimenters are effective expectation raisers for white children and black experimenters are effective with black children. White experimenters are effective with inner-city black children only in the case of third grade children. Black experimenters are effective with white children from both rural and suburban schools. The total number of children in these racial matching experiments is 472.

(2) In one of the few integrated elementary schools in Baltimore (approximately 60% black children and teachers), we performed the expectation raising experiment using groups of two white and two black children, and experimenters of both races. White experimenters were able to raise expectations for both black and white children drawn from mixed-racial groups of children; black experimenters, only with black children drawn

from the mixed-racial groups. We interpret the first finding to indicate that black children in this school, in contrast to black children discussed in the third experimental series, do not perceive white adults as hostile and insincere. The possibility exists that the second finding indicates that white children at this school do perceive black adults as hostile. The number of children in this set of experiments is 116.

A second finding for these studies of integrated groups is that in Phase I, before any intervention, we do not find evidence that children are forming expectations for themselves and each other on the basis of the race. This finding is striking in view of considerable research in segregated schools (for example, Cohen and Roper, 1972) showing effects of low expectations for blacks in mixed racial groups. We interpret this result to be one of the few quantitative pieces of evidence of positive consequences of school integration. (See detailed reference: Entwisle and Webster, 1974b.)

Fifth Experimental Series

In this set of experiments we addressed ways in which expectations might generalize. Do high expectations for one task spill over to affect other tasks? We noted earlier that often a teacher cannot honestly give positive evaluations to a student's early problem solving attempts at a new task. We therefore set out to explore what might be termed "contagion effects."

For this we developed a second task called "meal planning." Phase I and Phase III utilize the story-telling task, as before. Phase II uses only a meal planning task in which children select from a board pictures of foods which would go together well in a holiday meal. In these "generalization" experiments, we assume that high expectations for the meal planning activity will "generalize" to the story-telling activity.

Results of experiments with 80 white children and 72 black children confirm our prediction. In both sets of experiments the race of experimenters and children was matched. (See for detailed reference: Entwisle and Webster, 1974c (appended).)

The Evaluation Experiments

The five series using the basic experimental design described in the foregoing completed our investigation of several processes involved in the expectation-evaluation sequence as we hypothesized it to operate between teachers and children in classrooms. A different sort of consequence is also predicted to depend upon expectations: the nature of unit evaluations given performances. The higher the expectations held for a given actor, the more likely is anything he says to be positively evaluated. This process is represented in Figure 1; it is another crucial unobserved link in our analysis of the Pygmalion studies, since it models how teachers' expectations get transferred to students.

The evaluation experiments (n = 360) explored the evaluation process in three slightly different conditions. In all conditions, an entire class of students was given sentences supposedly completed by students at another school. Each sentence was described as a performance from a student either or high ability or of low ability. In one classroom, odd numbered sentences were described as coming from high ability students; in a second classroom, even numbered sentences were described as coming from high ability students. In condition 1, only the "ability" levels of sentence authors were given. In condition 2, "ability" was described, and this ability was said to be relevant to story-telling ability. In condition 3, the ability was described and said to be relevant, and in addition, the author was said to have told good (or bad) stories in the past. The three conditions thus are increasingly defined social situations. According to the theory, all should produce about the same degree of effect.

To measure expectations the children who are subjects give a "grade" to each sentence. The grade given when the author is described as having high ability is compared to the grade given when he is described as having low ability. The same sentences are used for both. In all three conditions of social specification, most sentences were judged as better when the author was described as having high ability than when the supposed author had low ability. The entire set of experiments was run twice, six classrooms per replication. Effects are entirely consistent over the two replications. (See for detailed reference: Webster and Entwisle, 1974 (appended).)

Summary

This completes the experimental research conducted under this program. Experiments have been devised to study the formation and the consequences of expectation states among elementary school children, and over 1,000 children have participated to date in these experiments. Figure 4 gives a summary of results and illustrates the major features and outcomes of these experiments schematically.

In a reconstruction of a research program, there is often an appearance of simple linear progression towards understanding and truth. What is more accurate, of course, is that investigators progressed <u>from</u> some point to the next point and then on from there. Throughout, our strategy has been to apply the perspective of expectation states theory to certain problems important in education, and we have moved from preliminary results and confirmation to new issues. The theory and its laboratory confirmation gave us a starting point and provided us a guide for our classroom research. Neither of these could foretell final results until the field experiments were done. The research program, as shown in the appended reports, contained numerous instances of unexpected findings or ideas which emerged as a result of the research process. These have been consistently followed up, and have often led to considerable increases in our understanding.

Figure 4
Summary of Experimental Program

	Experiment	Research Question	Ss
1.	Basic	task development, experimental design	gr. 3,4: black city; white rural
2.	Secondary	generality of phenomenon; sociological subgroups (sex, race, class)	gr. 1-4: white suburban
3a.	Special Control	"behavior modification"	gr. 3,4: white suburban
3b.	Racial Match	"hostility" vs. "debilitation"	gr. 3,4: black, city
3c.	White, Suburban, Female	contrast effect	gr. 3,female: white, suburban
4a.	Racial Mix	race of evaluator	gr. 1-4: white, rural; white, suburban; black, city
4b.	Mixed-Race Groups	effects of integration	gr. 3,4: integrated, city
5.	Generalization	raising expectations indirectly	gr. 3,4: white, rural; white suburban
6.	Grading	distribution of peer evaluations	gr. 3,4: white, suburban

We do not consider however, either that we are heading towards a specifiable goal, or in fact that there is a definable endpoint. The program is continuing, as indicated in Section II of this report. In addition to the main work and the minor issues raised during the research, we have explored and finally abandoned some other avenues. Three of these are (1) an experimental task, "artistic judgment ability," which did not meet criteria for the experiment; (2) the Draw a Man task employed in research by others (see Porter, 1971), which did not produce any reliable, interpretable results with our children; and (3) an experiment in which we hoped to demonstrate that children will allocate action opportunities to each other as a function of the expectation states they hold, which appeared to conflict with cultural norms against admitting differential evaluations of individuals. These topics are described more fully in Appendix H.

SECTION I

APPENDICES

Published Papers Covering Experiments on Children's Expectations

Reprinted from SOCIAL SCIENCE RESEARCH, Vol. 1, No. 2, June 1972

Social Science Research 1, 147-158 (1972)

Raising Children's Performance Expectations[1]

DORIS R. ENTWISLE and MURRAY WEBSTER, JR.

Department of Social Relations, The Johns Hopkins University, Baltimore, Maryland 21218

An experimental procedure to increase children's rate of hand-raising is reported. The procedure was tried with 240 grade school children from two residential areas and is apparently effective. This procedure, involving raising children's expectations for their own performance, differs in approach from previous work raising teachers' expectations. Similarities between the procedure and laboratory research in expectation theory are noted. The experimental results are useful in interpreting educational research designed to raise teacher's expectations and also useful in suggesting further work along similar lines.

Several investigators have reported that grade school children can be affected by the expectations held for them by their teachers (e.g., Rosenthal, 1968; Meichenbaum, Bowers, and Ross, 1969; Brophy and Good, 1970). One person's expectations of another's behavior may act like a self-fulfilling prophecy. Previous work in this area has not been systematic, however, in the sense of explicitly specifying the determinants, the definition, or the consequences of these performance expectations. For example, some of the studies cited point to the importance of expectations in educational settings without attempting to specify precise mechanisms at the individual level that alter expectations. In many of these studies teachers are given false information about students' test scores and then students subsequently perform in accordance with the false scores. We presume the teacher raises his expectations for the student when high (false) scores are given. Then this revised opinion causes the teacher to act so as to allow the student to participate

[1]This work was supported by NIH Grant MH-18183-01, OE Grant OEG-3-71-0122, and by the Center for Social Organization of Schools, The Johns Hopkins University. We thank Margaret Boeckmann for her help devising the experimental procedure, and Ellen Dickstein and Barbara Bricks for their help with the field work. Special thanks are due to Mr. Jack Epstein and Dr. George Gabriel for arranging for this research in schools. Principals who cooperated in the research helped in many ways. They are: Mr. Lyman Huff, Mr. William H. Magzis, Mr. Daniel Rickowiak, and Mr. Samuel Sharrow. Mr. Elliott Epstein and teachers and students of School 53, Baltimore City, were of great help to us in pilot studies preceding the experiments.

more, and more often to give positive evaluations of performances of that student. Once this chain of events gets started, there are powerful forces to cause it to continue. The student, perceiving the teacher's high expectations for him, tries to justify those expectations. More importantly, the student probably raises his expectations for his own future performance. With his expectations for his own performance raised, the student is more likely to participate in class, to receive favorable responses from others, to have a high opinion of his own ability, and the like. A central feature of this sort of explanation is the circularity of the process involved: raising the teacher's expectations for a given student causes the teacher to treat that student differently--to be more likely to call upon him in class and to evaluate positively anything he says. This in turn causes the student to raise his own expectations for himself, and thus to behave differently--to attempt to answer questions more frequently, to be more confident of his answers, and to expect (and actually to receive) more agreement from other students. If a teacher is given information which leads to the formation of high expectations for a particular student, it is likely that the teacher will treat the student in a way which will raise the student's expectations for his own performance. If this happens, the student will then behave in a way which will be positively evaluated by the teacher, and consequently in a way which will cause further raising of the teacher's expectations for the student. The expectations which are held *for* students, therefore, are extremely important determinants of future interaction and they are partially independent of "objective" evaluations of performance. Furthermore, expectations of a given level tend to be self-maintaining, for once they exist, they affect the very interaction components which determine expectation levels.

The main contribution of the present paper is reporting of an experimental procedure to raise directly children's expectations for their own performance and to do this without resorting to deception. Previous classroom research, as noted above, has attended to teachers' expectations, but children's expectations may be of equal, if not greater, educational significance. If expectations of children are to be altered in ways that are of use in the actual process of education, the alteration cannot rest on the giving of false information (as in the Rosenthal studies) or on the lowering of some children's expectations so that others' can be raised (as when children are randomly assigned to high or low tracks within a school). The procedure reported here seems to avoid these drawbacks.

The experimental task developed was a modification of a storytelling task previously used in research on cognitive development of children (Entwisle, Grafstein, Kervin, and Rivkin, 1970). In addition to meeting the major criteria listed above, it also meets practical criteria: (1) it is similar to activities that ordinarily occur in classrooms; (2) it provides discrete, easily observable performance outputs that permit clear evaluations; (3) it is interesting enough to capture children's attention.

METHOD

Subjects. Children were drawn from four schools in the Baltimore area, two inner-city schools with nearly 100% black students, and two rural schools with nearly 100% white students. The rural schools are located in a farming district 30 miles north of Baltimore, near the Pennsylvania line. For the two inner-city schools, only students whose school records showed tested IQ scores between 90 and 110 were selected for the study. The IQ range for rural students is considerably larger, from 76 to 141, but the mean IQ is 105 for third-graders and 108 for fourth-graders. Experimenters were middle class white persons in the 20-30 age range, both sexes.

Procedure. At the beginning of an experimental session, children were brought together and told that the researchers were looking for people who could tell good stories. They were to be divided into "teams" and were told that the team which made the best stories would win a prize. Then one experimenter took the members of each team to a separate room and described the storytelling task to them. Members of a team were chosen so that children on one team came from different classrooms.

The experiment has three phases. In Phase I, the "baseline" level of volunteering (hand-raising) is determined for all children; in Phase II, the attempt was made to manipulate upward the self-expectations of the child selected as the "experimental" child; and in Phase III, the rate of volunteering (hand-raising) was again measured. The Phases are summarized in Table 1.[2] Children were told that the "game" consisted of making up a story. The experimenter would help by starting sentences, but then the children should try to make interesting stories by supplying "good" words when asked for words for the story. In each instance the same story skeleton (see Chart 1) was used, and the skeleton was missing 12 words. Children supplied these 12 words.

Children were told to listen carefully while the sentence was being read, then when the blank was reached, to try to think of a good word. Anyone who thought of a good word was to raise his hand, and the experimenter would select one child to give the "team's word" for that sentence. Children were cautioned not to raise their hands unless they thought they had a good word, for if they were called on and gave a "bad" word, this would hurt the team scores. The experimenter allowed 30 sec to elapse after reading the sentence before calling upon a child.

[2]A few children in the inner city schools told two stories in each phase. Only the first story in each phase was evaluated to make the procedure as consistent as possible with the procedure followed for most of the inner city children and for all of the rural children. Also in a few instances for inner city children there were five rather than three control group subjects in an experiment. In such cases two subjects per group have been discarded using a random procedure.

TABLE 1

Summary of Experimental Procedure

	Phase I	Phase II	Phase III
Control *Ss*	One story is produced (12 words); no evaluations; level of volunteering observed	Control *Ss* have story read to them by another adult; from 12 to 15 control *Ss* join in one group	Repeat Phase I, with same control *Ss* and experimental *Ss* as in Phase I; experimenters are rotated so the experimenter is unaware of identity of experimental *Ss*
Experimental *Ss*	One story is produced (12 words); no evaluations; level of volunteering observed	Experimental *Ss* make up story individually with the same experimenter they have seen in Phase I; receive all positive evaluations; experimental *Ss* join control *Ss* at end	Repeat Phase I, with same control *Ss* and experimental *Ss* as in Phase I; experimenters are rotated so the experimenter is unaware of identity of experimental *Ss*

CHART I

Story Skeleton with Sample Entries from a Rural Group

There was once a (very tall prince)

Who had a (castle)

That (he, she) (lived in)

One day (he, she) had to go to (the dungeon to see his prisoners)

(He, she) did this very (angrily)

Because (he, she) wanted to (make sure they were there)

This was very dangerous because of the (strong prisoners)

Who (which) was (were) very (mean)

In order to fool the (fill in) the (fill in) dressed up

As a (another prisoner)

In such a disguise the (fill in) looked (mean)

And when the (fill in) saw the (fill in), they (welcomed him)

This made the (fill in) (feel pretty good)

And /complete story/ (he let his new friends go.)

RAISING EXPECTATIONS

Before calling on a child, the experimenter recorded privately which children were holding their hands up (the measure used for the child's expectation state). The experimenter held a clipboard so that children could not see what was on it. The clipboard was used for recording words given by the children. While the 30-sec waiting period was occurring, it was easy for him to make small marks indicating which children had volunteered without the students' being aware of his recording.

Phase I consists of an initial story being produced as just described. The experimenter did not evaluate any of the words given during Phase I. He called upon each child for Phase I approximately the same number of times. With 12 words to supply and four children playing, this permits each child to be called upon three times. In only a few instances did any child volunteer less than three times altogether and so unbalance the selection of respondents.

At the end of Phase I children were told that they should now return to the room where they had initially assembled. After they began to move out, the experimenter said quietly to one child (chosen because his level of hand-raising was near the median for the group) that he should stay in the experimental room and wait there for a minute for the experimenter. After the experimenter made sure that the control children were on their way to the proper destination, he returned to the room and to the selected (experimental group) child. The experimenter then told the experimental child that he would have an opportunity to make up a story all by himself; also that the experimenter had played with many children making up stories and that the experimenter thought he (the experimental child) was really very good at the task. Then a story-skeleton with a new lead word was filled in orally by the child, as before, but of course there was *no* volunteering. The child merely supplied a word when the sentence was read up to the blank. After each word was supplied, the experimenter indicated approval vigorously –by smiling, by nodding, by commenting a "very good word"–"good"–"that's interesting!", etc.; i.e., he indicated approval in every way possible consistent with sincerity.

When Phase II storytelling was completed the experimental child went and joined the control children in Phase II. Phase II for control group children was a storyreading session. Children from several game groups gathered in a central room as they finished Phase I. The storyreading went on continuously and experimental group children also joined the storyreading group as they finished their Phase II activities. The experimental children thus listened to the end of the story being read to the control groups. The experimental children's entrance was not noticeable because they were part of a large group where other members had also arrived at different times. The storyreading prevented communication among the children in Phase II. At the end of the storyreading, the children were told to "go back to the room where you were before." *All* children thus returned together. At this point experimenters were

rotated among rooms so each experimenter had a new group and was unaware of the identity of the experimental child. The length of each phase varied with the Phase II control procedure adjusting its length to the needs of each set of experiments.

Phase III, except for the rotation of experimenters, was a repetition of Phase I. The four children made up a story; the experimenter remained neutral and recorded unobtrusively how many times each child raised his hand.

RESULTS

The results of the experiment are given in Tables 2 and 3.[3] The data consist of gain scores for each child, the increase in the number of times each child raises his hand in Phase III compared to the number of times he has raised it in Phase I. The differences between experimental and control group means are as predicted, and increases in average volunteering rate vary from 0.97 to 2.03. The analysis of variance in Table 3 shows that this difference in response to treatment is highly significant. None of the interactions is significant; the response to treatment does not vary significantly by grade or residential locus.

DISCUSSION

Results suggest that the experimental procedure devised does lead to changes in rates of volunteering, which, we believe, represent changes in

TABLE 2

Average Gain in Rate of Volunteering from Phase I to Phase III For Experimental and Control Groups (*N*s are in Parentheses)

	Inner city		Rural	
	Grade 3	Grade 4	Grade 3	Grade 4
Experimental group	1.85	1.90	2.60	1.90
	(20)	(10)	(20)	(10)
Control group	0.88	0.20	0.57	−0.07
	(60)	(30)	(60)	(30)
Difference between groups	0.97	1.70	2.03	1.97

[3]Altogether 67 observations were discarded to yield proportional subclass numbers. Each subclass is approximately balanced for sex. An earlier least-squares analysis of the unbalanced data is consistent with findings reported here. Strict additivity holds for the case of proportional subclass numbers, and as in the present analysis, all effects are orthogonal. Bert Green kindly supplied a proof for the orthogonality of proportional subclass numbers for three variables.

TABLE 3

Variance Analysis: Gains in Rate of Volunteering, Rural and Inner City Children, Grades 3 and 4

Source of variance	*df*	Sum of Squares	Mean square	*F* value
Experimental vs. control treatment (T)	1	116.81	116.81	18.19[b]
Residential locus (R)	1	0.60	0.60	0.09
Grade (G)	1	17.63	17.63	2.75
T × R	1	7.20	7.20	1.12
T × G	1	1.12	1.12	0.17
R × G	1	0.30	0.30	0.05
T × R × G	1	1.58	1.58	0.25
Within T-R-G groups	230[a]	1476.72	6.42	

[a]Two missing observations were estimated using appropriate subclass means.
[b]Beyond the 0.01 level.

self-expectation states. In no previous work has there been an attempt to alter children's expectations directly, and in all previous work there has been major reliance upon the use of false information. While techniques using false information may be very effective and consequently useful for research purposes (especially in laboratory studies where the deception involved may be explained immediately afterward), false information is clearly not desirable for long-term use in applied research. In studies where expectations are raised, there are practical difficulties associated with the continued use of false information; in studies where expectations are relatively lowered, as in some tracking studies, there are moral difficulties. The maneuver we have described involves an adult interacting with a child on an individual basis. The adult gives consistent positive evaluations of performance in a task where the child's actual ability is almost irrelevant. The evaluations are therefore not inconsistent with anything known about the child or with his potential. To the extent that the maneuver is successful in changing the child's behavior, it should improve his performance at a wide-variety of tasks.

The nature of the experimental maneuver. To clarify the nature of the experiments reported here it is important to define exactly the experimental procedure so as to rule out rival explanations for the results.

Maneuvers that give fictitious results to students or to teachers may be defended when they are one-shot procedures to demonstrate self-fulfilling prophecies, placebo effects, and the like, or when they are needed to allow quick and sizable manipulation of variables in the laboratory. Other methods must be sought, however, when the goal is to change children's long-range expectations for themselves, or others' expectations for children in classroom settings. If expectation-alteration is seen as one goal of education, ways must be found to alter expectations that can be defended over the long-term and

that are compatible with other educational goals. The procedure described here has achieved some modest success along these lines. One issue, that of biasing of results, is particularly salient, for it is well-known that even a slight change in manner of a teacher can alter significantly the likelihood that a child will raise his hand. A number of steps regularly were therefore taken to minimize the effects of any systematic influence from the experimenter. First, in Phase I of these experiments, *E*s were trained to treat the children as equally as possible and to refrain from evaluating any of their performances. The child selected as the "experimental" child at the end of Phase I was chosen on the basis of his having responded close to the median for his group. To choose a child with too low a response rate in Phase I might have biased results in favor of predictions through a "regression to the mean" phenomenon in Phase III. To choose a child with too high a response rate in Phase I would have biased results against predictions, because of a "ceiling effect" in the measure of expectation states used. To avoid drawing attention to the child selected, the *E* did not say anything to that child until the others were on their way to the other room for Phase II; nor did the *E* explain at all to the others why this child was asked to remain behind. *All* children are with an experimenter during Phase II, and the experimental group children participate in at least part of the control treatment of this phase.

The major danger of biasing would come in Phase III, however, if the *E* responded more warmly or more positively to the "experimental" child. In addition, this child might feel that, because of the individual session with the *E* in Phase II, he had some "special" relationships with him.[4] In order to avoid problems of this nature, *E*s were rotated before the beginning of Phase III. Thus the *E* was new to the group in Phase III, unfamiliar to both the experimental and control children. Furthermore, the *E* did not know at this point which child had been given the Phase II "treatment," and thus was unlikely to treat the children differentially. In Phase III, as in Phase I, *E*s are trained to call on every child as nearly as possible the same number of times.

Probably the major alternative interpretation of these results is one which would describe the Phase II procedure as *reinforcement* of behavior rather than *positive evaluation*. In other words, a major alternative explanation for these results would not use the idea of expectations at all, but would argue instead that the experiment demonstrates that childen will increase the rate of emission of behaviors which have been rewarded in the past. Interpreting the results of this study in terms of behavior reinforcement would mean arguing that there is no reason to believe that the experimental group

[4]A later series of experiments (forthcoming) with additional control groups rules out explanations based upon isolation with the experimenter. This result has also been found in highly controlled laboratory studies by Professor Barbara Sobieszek at the University of Rochester. We thank Professor Sobieszek for making her data available to us prior to publication.

children changed their conceptions of their own abilities, or that they would be more likely to display the behavioral consequences of a raised self-concept, such as increased confidence and willingness to try to perform in other situations. In assessing such an explanation, it is important to note that our measure of expectation states in Phases I and II was rate of volunteering, or *hand-raising*. What were positively evaluated in Phase II were *words spoken* by the children. At no time during the experiment was anyone positively evaluated (or reinforced) for hand-raising. Furthermore, the specific words evaluated in Phase II were seldom the words volunteered in Phase III. Thus, an explanation based on the child's expectations is more parsimonious than one based on a reinforcement paradigm.

Additional support for the interpretation given our experiments is found in experiments which measure children's expectations by using a kind of self-rating scale. They consistently report that expectations are increased after social reinforcement by an adult, and are unchanged when an adult maintains a neutral role (Hill and Dusek, 1969; Crandall, 1963; Crandall, Good, and Crandall, 1964). The actual procedure used by Hill and Dusek was very similar to our own, for an adult responded "That's good. Fine. Very good. You're doing well" for positive reinforcement following attempts at an angle-matching task. The adult was neutral and nonresponsive in their nonreinforcement condition (like our Phases I and III procedures). Other studies (Maehr, Mensing, and Nafzger, 1962; Videbeck, 1960) showing that if an individual gets approving reactions from others with respect to some specified attribute he will improve his self-rating on that attribute, are also consistent with the conclusion that children's expectations have been improved by our experimental maneuver.

With respect to alternative interpretations of the results of the study, obviously we cannot rule out all rival explanations with perfect certainty. However it seems difficult to construct an explanation which accounts for *all* the results of this study, as well as for the results of others' work mentioned earlier, as satisfactorily as does an "expectation" interpretation. For example, it might be argued that the effect of the treatment of the experimental group child in Phase II was to give him additional practice with the task of storytelling, and that it was the effect of the practice which produced the increased confidence in his ability to tell stories. This would mean that the positive evaluations of performance in Phase II were unimportant, and possibly that the level of expectations as determinants of behavior in Phase III were irrelevant as well. But it is important to note that the task is one which calls for no special ability, and certainly not one which can be "improved" through practice In the absence of performance evaluations from *E*, it is probably difficult for children to decide whether even their own words are "good ones"; the subjective impressions of the experimenters in the "unevaluated" Phase I corroborate this.

Relation to previous research. Some research to which the present study is closely related is laboratory work in expectation theory by Berger and his associates (1961, 1966, 1968, 1969). Expectation theory assumes that in task situations positive expectations are communicated by a source (an experimenter or an evaluator) who gives positive or negative evaluations of performances. Another assumption is that the person receiving the evaluations comes to hold an expectation state that is in accord with those evaluations. One can then derive the proposition that if one person receives a large number of positive evaluations (experimental group children) while a second person has not received any evaluations (control group children) then the first individual is more likely to accept an action opportunity (raise his hand) than the second person. The increased acceptance of action opportunities would be seen as an improved state of self-expectation or of self-evaluation. The present experiment operationalizes variables in a way similar to the way such variables are operationalized in the laboratory. The present experiments in naturalistic settings are entirely consistent with this laboratory work. (The current work of Cohen and associates (1970) involves applying the same theoretical perspective to a different set of substantive problems.)

Other recent work analyzing behavioral concomitants of expectancy is also consistent with both the laboratory work and our work. For example, Meichenbaum *et al.* observe that expectancy instructions (identification of "late bloomers") cause some teachers to increase positive interactions with students, or to decrease negative interactions. Positive interactions included conveying encouragement, praise, or any attitude of satisfaction. Also Brophy and Good (1970) observe that first grade teachers tend to demand more of those children for whom they hold high expectations, and to praise these children more when they do respond. Students seen as low are praised less and held to lower standards. In other instances when expectancy or teacher expectancy experiments have failed to achieve results that were anticipated (Goldsmith, 1970; Fleming and Anttonen, 1970) an analysis of findings in the light of the laboratory work of expectation theory may be helpful. For example, it may be that the teacher does not alter expectancy because students do not accept the validity of the teacher's expectations.

Educational implications. The major result of this work is the experimental task itself. It offers a means of manipulating (raising) young children's expectations that is not subject to the drawbacks affecting ways of manipulating expectations used previously, and even more important, the raising of the expectations of the child himself is probably a more direct and more general way of improving educational performance than the changing of other persons' expectations for the child. The behavior influenced-frequency of hand raising is generally considered to be important educationally. Hand-raising leads to greater participation. The importance of active participation to learning is too well-known to require documentation. In future reports we

hope to present the results of studies of the degree of effect of this procedure upon racial, age, and SES subgroups of the population. Further work will extend the basic experimental design to examine the relation between expectation change and structural variables like sociometric standing.

The present experiment includes children from grades 3 and 4, from two very different residential loci. Since there are no significant differences by locus or by grade but a highly significant gain in rate of volunteering when experimental children are compared to control group children, it seems likely that the treatment devised is an effective one for many children of this age level.

Assuming future work agrees with results reported here, one must evaluate the potential of this procedure for educational purposes. There is value first of all, in making explicit parallels between social psychological research and research in classrooms. As has been so frequently noted, there is a surprising gap between the two fields of research that hinders both. By linking classroom research to more tightly controlled research, one gains coherence and explanatory leverage on a whole body of research findings.

The viability of the experimental maneuver reported here for securing long-term educational effects may not be large, however. The present procedure, as pointed out earlier, does have the distinct advantage of avoiding the use of false test scores or of any form of outright deception. On the other hand, it is difficult to imagine using such a maneuver over and over to obtain general effects in raising children's expectations. Yet the procedure may have usefulness over the short term in two ways: (1) The procedure may act as a pump primer. That is, if a child's expectations for himself are suddenly raised as in the experiment, he may alter his actions in ways consistent with his increased expectations for himself. Any intervention within the circular series of events—improved self-expectations leading to better performance leading to improved teacher evaluations—may be effective. The short-term change in a child's actions induced by an expectation-raising procedure might thus fire a chain reaction that would tend to continue. (2) This procedure and others that could be invented may give precise suggestions as to how teachers can convey positive expectations. Some teachers may wish to convey positive expectations but not be very adept at it. The suggestion here is that encouragement for one activity (word-giving) leads to an increased level of another activity (hand-raising) at a later date. In classrooms where many activities occur, then, the teacher need not wait, perhaps in vain, for a praiseworthy performance in arithmetic to improve a child's expectations for himself in the area of arithmetic. If the child can be reinforced for performance of another kind, this may generalize to yield increased participation across the board. In fact, some situations may be used primarily to improve expectations rather than for learning *per se*.

REFERENCES

Berger, J., Cohen, B. P., and Zelditch, M. (1966), Status characteristics and expectation states: a process model, *in* "Sociological Theories in Progress" (J. Berger, M. Zelditch, Jr., and B. Anderson, Eds.), pp. 47-73, Houghton Mifflin, Boston.

Berger, J., and Conner, T. L. (1969), Performance expectation and behavior in small groups, *Acta Sociologica* 12, Winter.

Berger, J., and Snell, J. L. (1961), A stochastic theory for self-other expectations, Technical Report No. 1, Laboratory for Social Research, Stanford University, Stanford, California.

Berger, J., Zelditch, Jr., M., Anderson, B., and Cohen, B. P. (1968), Status conditions of self-evaluation, Technical Report No. 27, Laboratory for Social Research, Stanford University, Stanford, California.

Brophy, J. E., and Good, T. L. (1970), Teachers communication of differential expectations for children's classroom performance, *The Journal of Educational Psychology* 61, 367-374.

Cohen, E. G., Lohmann, M., Hall, K., Lucero, D., and Roper, S. (1970), Expectation training I: altering the effects of a racial status characteristic, Technical Report No. 2., School of Education, Stanford University, Palo Alto, California.

Crandall, V. C. (1963), National Society for Study of Education.

Crandall, V. C., Good, S., and Crandall, V. J. (1964), Reinforcing effects of adult reactions and nonreactions on children's achievement expectations: a replication study, *Child Development* 35, 485-497.

Entwisle, D. R., Grafstein, D., Kervin, J., and Rivkin, M. (1970), Giant steps: a game to enhance semantic development of verbs, Report No. 81, Center for the Social Organization of Schools, The Johns Hopkins University.

Fleming, E. S., and Anttonen, R. G. (1970), Teacher expectancy or my fair lady, Unpublished paper presented at the annual meetings of the American Educational Research Association.

Goldsmith, J. S. (1970), The effect of a high expectancy prediction on reading achievement and IQ of students in Grade 10, Unpublished M.Ed. Thesis, Rutgers University.

Hill, K. T., and Dusek, J. B. (1969), Children's achievement expectations as a function of social reinforcement, sex of S and test anxiety, *Child Development* 40, 547-557.

Maehr, M. L., Mensing, J., and Nafzger, S. E. (1962), Concept of self and the reaction of others, *Sociometry* 25 353-357.

Meichenbaum, D. H., Bowers, K. S., and Ross, R. R. (1969), A behavioral analysis of teachers expectancy effect, *Journal of Personality and Social Psychology* 13, 306-316.

Rosenthal, R., and Jacobson, L. (1968), "Pygmalion in the classroom," Holt, Rinehart, and Winston, New York.

Videbeck, R. (1960), Self-conception and the reaction of others, *Sociometry* 23, 351-359.

Research Notes: Status Factors in Expectation Raising*

Doris R. Entwisle
Murray Webster, Jr.
The Johns Hopkins University

Previous experiments with grade school children, designed to raise their expectations for their own performance, were repeated with pupil subgroups formed according to race, sex, and age. Positive results of the earlier studies were reproduced, and further analyses are presented to assess the effect of some status factors in the situation. The experimental procedure was effective across the subgroups studied. The relation of this work to "teacher expectancy" studies is discussed, and some implications of the results, in terms of other studies on expectations and the structure of competition in schools, are drawn.

PERFORMANCE EXPECTATIONS, LIKE other beliefs individuals hold about themselves, can act to produce a self-fulfilling prophecy. At the most general level, people who expect success in a given task are often more likely to meet success than are people who expect to fail. Behaviors such as sustained attempts to solve a problem, expressions of self-confidence, frequent verbal interaction, and refusal to accept influence from others are thought to be determined, in part, by an individual's expectation level and, in turn, to be partial determinants of his actual success.

Expectations for success are of continuing interest in education. For a long time it has been felt that the child's expectations for his own success and the expectations which are held by significant adults (such as teachers or parents) for the child's success are important determinants of his actual performance.

* Much of this research was supported by Grant No. MH 18183. Preparation was facilitated by Grant OEG-3-71-0122. Special thanks are due to Dr. Evart Cornell, principal, and the teachers and students at Stoneleigh Elementary School for their help in these experiments.

Sociology of Education 1973, Vol. 46 (Winter): 115-126

Expectations, either as the major independent variable or as an intervening variable, have been invoked as the explanation for a large number of results involving children's academic performance. (See Meichenbaum, Bowers, and Ross, 1969, for a recent review of much of the related literature.)

Some investigators have demonstrated positive effects by artificially raising expectations held for specific children (for example, Rosenthal and Jacobson, 1969), but such research has produced contradictory and unexpected findings as well (for example, Claiborn, 1969). The inconsistencies in expectation research are perplexing. This research has not been guided by explicit theory specifying how expectations may be raised or the conditions under which various treatments will be successful or unsuccessful. Typically there has been neither any way of knowing exactly why a given procedure should work nor of interpreting negative results when a procedure did not work as predicted. The absence of explicit theory may be an important reason for the mixed results.

The child in the classroom is in a situation similar to that studied in highly controlled laboratory experiments on *expectation theory* by Berger and his associates (Berger and Snell, 1961; Berger and Connor, 1969; Berger, et al., 1969). The child is acting as a member of a problem solving group, he usually is motivated to do well, and he perceives that the task of the group as a whole is to solve a series of problems. Because of these similarities, we felt that research attempting to raise children's expectations could be fashioned along the same lines as laboratory experiments in expectation theory. (See Entwisle and Webster, 1972, or Webster, 1969, for a statement of the propositions of the theory)

We designed a study based upon the propositions of expectation theory, devising procedures that proved to be successful in raising children's expectations for their own performance. In an earlier report (Entwisle and Webster, 1972) we discussed task development, some initial experiments, and the relation between educational field settings and laboratory settings. This first work was essentially exploratory and psychological in nature: we were concerned with developing effective means of applying the theory to the learning situation and with determining whether we could affect the child's cognitive expectations for his own performance. Here we report results from further work designed to examine the effect of sociological variables: whether, for example, the results of the earlier study could be replicated using a different sample and across variations of age, race, and SES subgroups of the school population.

To review briefly the earlier study, we devised a three-phase experimental treatment to increase the self-expectations of children

and we measured the child's expectation level by the rate of acceptance of action opportunities, that is, hand-raising in response to questions. Children were formed into groups of four; each group met with a single experimenter. They were told that their task was to help make up stories and their group was in competition with other groups to make up the "best" or most interesting stories. The children were asked to supply words to fill in blanks in a "story skeleton" consisting of 12 sentences with one word missing from each sentence.

The invention of each of three stories constituted a "phase" of the experiment. In phase I, the experimenter was neutral: he recorded the proportion of times each child raised his hand and called equally on children who volunteered to supply the 12 missing words. In phase II, one child constructed a story by himself (experimental group), and the other children (control group) went to another room where they played a game or had a story read to them. Every word the experimental group child gave in phase II was evaluated positively by the experimenter; according to the theory, this procedure should have raised the child's expectations for himself and, consequently, his rate of hand-raising. In phase III, the original groups were reconstituted, experimenters were reassigned to groups to prevent unequal treatment of the experimental group child, and the measurement procedure of phase I was repeated.

In the initial studies, the experimental group child in phase II generally increased significantly the rate at which he volunteered words between phase I and phase III; the increase was significantly greater for experimental children than for "control" children. Thus, results of the initial studies were encouraging—the procedure and the theory seemed to work as predicted. The research included two groups of third and fourth grade students. One group was black and lives in the inner city of Baltimore; another group was white and lived in rural Maryland. These groups were diverse in many respects such as life style and family environment, but in other ways they resembled one another—for example, in economic level and educational problems.

This paper reports an extension of the earlier research in two important directions: first, by studying white, middle-class, suburban children, the social class or subcultural group dimension was extended; second, by studying children from first and second grades, as well as grades 3 and 4, the age and amount-of-schooling dimension was extended. These extensions when combined with the earlier work now provide data for third and fourth grade children from three widely different cultural backgrounds (black ghetto, white suburban, white rural) and give some notion of how younger children (grades 1 and 2) react.

The studies reported here are based on a three-phase experimental design, where the second phase included a procedure to raise expectations of a randomly selected child. (The reader is referred to Entwisle and Webster, 1972, for complete description of the design.) Subjects included approximately equal numbers of boys and girls, with 79 first graders, 84 second graders, 112 third graders, and 103 fourth graders. All members of the first four grades in a single suburban school participated. Some groups were later eliminated randomly to produce proportional numbers of children in the various subclasses. Each story-writing team, as before, consisted of members of a single sex and grade. Insofar as possible, the four members of each team were chosen from different classrooms. For some grades that were distributed among three classrooms, two children were taken from a single classroom and the remaining two from two other classrooms.

Results

Table 1 shows the mean gain in rate of volunteering from phase I to phase III for both the experimental and the control groups. These groups are balanced approximately for sex. Three sorts of changes in rate of volunteering are reflected in Table 1. First, all groups, control as well as experimental, showed some increase in rate of volunteering. An increase for the control groups could have been due to such things as increasing familiarity with the situation and excitement over the competition. Second, for all grades, the increase was significantly greater for children in the experimental groups than for the control groups (see Table 2). The increase in the experimental groups is consistent with the predictions of the theory, and it constitutes a replication of the earlier finding with a different subject population. Third, second grade children showed the greatest absolute amount of increase for the experimental group and also the greatest difference be-

TABLE 1

Average Gain in Rate of Volunteering From Phase I to Phase III, White Middle Class Students

	Grade				
	1	2	3	4	All Grades
Experimental Group	1.20	2.55	1.85	1.85	1.86
	(20)	(20)	(20)	(20)	(80)
Control Group	0.67[a]	0.67	1.20	1.38[a]	0.98
	(59)	(60)	(60)	(59)	(238)
Difference between Groups	0.53	1.88	0.65	0.37	0.88

[a] This mean is based on 59 rather than 60 observations.

TABLE 2

Analysis of Variance for Gain in Rate of Volunteering, White Middle-Class Students Approximately Balanced for Sex

Source	d.f.	Sum of Squares	Mean Square	F-value
Experimental vs. Control Treatment	1	46.81	46.81	6.02[a]
Grade	3	22.42	7.47	0.96
Treatment x Grade	3	20.27	6.76	0.87
Within Treatment x Grade Groups	310[b]	2409.70	7.77	

[a] Beyond the 5% level.
[b] Two observations are missing, estimated by subclass means.

tween experimental and control groups in this study, but the grade x treatment interaction is not significant. The results of this study are consistent with the results of the previous study. Thus, the effects of increased expectations have been demonstrated for a considerably wider sample than previously.

The results presented in Tables 1 and 2 suggest that the experimental procedure has been successful in raising children's expectations as predicted. Children in the experimental groups increased their rate of volunteering between phase I and phase III significantly more than did children in the control groups. Putting these data together with results reported for the previous samples, we conclude that the expectation-raising procedure is effective for a variety of subgroups in the population: white rural children, black inner-city children, and white suburban children.

In Tables 3 and 4 further analyses are presented that combine data of third and fourth grade children in the present study with data procured in the previous study. (The data for the earlier samples differ slightly from those reported originally—Entwisle and Webster, 1972—in that data have been randomly discarded from some groups to provide proportional subclass numbers. The exclusions do not introduce changes in any results reported earlier.) Grade and treatment again are factors, and residential locus may now be added as a third factor. Half of the data from fourth grade children reported in Tables 1 and 2 have been eliminated randomly so that data for white suburban children when combined with earlier data will yield a matrix with proportional sub-class numbers.

Table 3 shows that the treatment was effective for all subgroups of third and fourth grade children in both studies. With the exception of rural fourth grade children, all control groups showed an increase, but for no subgroup was the increase for control group children as great as the increase for experimental group children.

TABLE 3

Average Gain in Rate of Volunteering From Phase I to Phase III, for Children from Three Residential Loci, Grades 3 and 4

	Suburban		Inner-City		Rural	
	Grade 3	Grade 4	Grade 3	Grade 4	Grade 3	Grade 4
Experimental Group	1.85	2.20	1.85	1.90	2.60	1.90
	(20)	(10)	(20)	(10)	(20)	(10)
Control Group	1.20	2.07	0.97	0.20	0.57	—0.07
	(60)	(30)	(60)	(30)	(60)	(30)
Difference Between Groups	0.65	0.13	0.88	1.70	2.03	1.97

Table 4 shows that the overall effect of treatment was highly significant ($p<.01$). There is no firm evidence here of differences by grade or by residential locus (residential locus approaches significance, $p<.10$), nor is any interaction of treatment with other factors significant. As in the earlier report, it seems fair to conclude that expectations were raised significantly in experimental group children compared to control group children. In addition, from these later studies we conclude that residential locus or grades does not significantly affect susceptibility to treatment because the treatment factor does not interact with these other factors.

We also ran a short series of more elaborate experiments where an additional control group was formed. In Phase II of this series, one of the control children met with a second experimenter, produced a story, and in every way received identical treatment to the experimental child's Phase II treatment except

TABLE 4

Analysis of Variance for Gain in Rate of Volunteering, Three Residential Loci, Two Grades

Source	d.f.	Sum of Squares	Mean Square	F-value
Treatment (T)	1	110.21	110.21	14.75[a]
Residential Locus (R)	2	43.62	21.81	2.92
Grade (G)	1	0.87	0.87	0.12
T x R	2	22.21	11.10	1.49
T x G	1	0.50	0.50	0.07
R x G	2	35.77	17.88	2.39
T x R x G	2	2.27	1.14	0.15
Within T, R, G Subgroups	347[b]	2593.15	7.47	

[a] Beyond the 1% level.

[b] Two observations are missing; estimated by subclass means.

that the experimenter remained neutral, not giving positive evaluations. Two other control children received the control treatment described earlier. It turned out that the gain in Phase III for these special control children did not differ from the gain for the usual control children, and that the gain for the experimental children significantly exceeded the gain for both control groups. Since this more elaborate procedure leads to the same outcome, its continued use is unnecessary. It is a much more difficult procedure to implement since it requires four experimenters for each group of four children whereas the less elaborate procedure used in most of our experiments can deal with three groups of four children simultaneously with the same number of experimenters.

Discussion

The results of this study extend the positive findings of our earlier study and demonstrate that white middle class children respond to the experimental procedure by increasing the rate at which they raise their hands. Although there are other subgroups that could be studied (for instance, black middle class) the range of children for whom the experimental procedure has produced increases in performance outputs suggests that the phenomenon probably is a general one. There are obvious similarities between the completion of a story under the experimenter's direction, the task in these experiments, and many tasks overseen by the teacher in the classroom. In at least a preliminary way, the scope of application of expectation theory now has been extended to situations that frequently exist in natural settings: specifically, elementary education for children of various subgroups.

In terms of the goal of modifying expectations in naturalistic settings, it is important to point out how the significance of social status in this study differs from its significance in other work related to expectation theory. Generally, studies of status characteristics and expectation states concern the effect of differential status in assignment of power and prestige within a small group. Thus, Cohen (1968, 1970) has studied the task performance of mixed groups of black and white boys who attempt to solve a problem together, and Webster (1970) has studied the effects of status characteristics on the effectiveness of an evaluator. In both Cohen's and Webster's studies, the significance of the status characteristic is that, under certain circumstances, *differential* conceptions of ability arise in accord with the different states of the diffuse status characteristics.

The focus of the present study differs. Here we have sub-

groups, all of whose members are *equal* with respect to a status characteristic (such as sex, age, and race) and our interest is whether the same experimental treatment will raise expectations equivalently for the various status groups represented in the study. The intent, thus, is to look at the "demography" of expectation ability, particularly for naturally-occurring social status groups of kinds typically attending elementary school. Elementary schools, since they draw from surrounding neighborhoods, and usually are not large tend to have students who are of about the same social class level. It appears that all children so far studied are responsive to this experimental treatment for modifying their expectations; no significant differences by residential locus have appeared.

Age level (over 4 grades) also was systematically varied and likewise may be viewed as a demographic variable. We do not find that the effect varies with age, but further study is necessary before age can be discounted.

In the analyses reported here we have not observed significant effects of either the age/school grade factor or the residence/race/SES factor. These negative findings should not be interpreted to mean that these factors are unimportant, either in affecting the expectations children have for their own performance or in attempts such as ours to raise these expectations. We might reasonably expect that the differential life experiences which are reflected in the factors of our analyses will produce differences in both the general level of performance expectations and in susceptibility to different sorts of treatments. However, what seems most important about the failure to find significant effects of these factors in our research is that, *by comparison with the experimental procedure used to raise expectations,* the social status differences are relatively unimportant. In other words, based upon the work so far completed, it seems safe to conclude that the procedures developed for this situation are effective quite generally for children coming from very diverse social backgrounds.

RELATION TO TEACHER EXPECTANCY RESEARCH. Results of teacher expectancy studies have been disputed in some cases (see Barber and Silver, 1968; Thorndike, 1969) and unequivocally negative in others (Claiborn, 1969; Jacobs, 1969; Fleming and Anttonen, 1970). Our experiments point to possible sources of difficulty in the teacher expectancy work. First, while most children show slight increase in performance outputs even without much encouragement (the control groups), some individual children fail to manifest increases in performance even with large amounts of positive evaluation. We have no rationale to explain this, since all experimental children were given the same treatment. Different experimenters, however, could produce different effects on children even when the actions and speech of an ex-

perimenter are the same from child to child. One child, for example, might have received much praise in the recent past while another might be relatively deprived with respect to praise. We then would expect the latter child to be particularly receptive to praise by the experimenter: the same remarks would be falling on two very different sets of ears. The more variability of this kind that there is among children, the more difficult it will be to demonstrate a significant effect. There probably are subtle kinds of teacher characteristics that decrease teachers' effectiveness as purveyors of expectancy effects, much as presumed subtle effects may have operated to cause differences in our experimental treatment.

Second, expectation theory assumes a task orientation on the part of students; it also assumes a task in which ability is equal or irrelevant. Both assumptions may be violated often in the teacher expectancy work, for many school children have little ego-involvement in academic pursuits and many already have firm ideas about their own supposedly low level of ability. Also, teachers' expectations are not the only expectations that are relevant. Children have some access to objective standards and to alternative others for evaluating their performances—parents, peers, principals—so the teachers' expectations for them may not be only one component in an overall set of expectation components.

A third point made by Claiborn (1969) is that there may be no changes in teacher-pupil interaction, no classroom analogue of our phase II treatment. If teachers perceive pupils to be of high potential, teachers may alter their behavior, but not all teachers do so (see Kranz, Weber, and Fishell, 1970), and teachers vary in how they change their behavior to suit children's ability. For example, Kranz, et. al. (1970) show that some teachers behave similarly towards high and average ability children but differently towards low ability children. Other teachers change their behavior towards high ability children and manifest similar behaviors towards average or low children. If, as in most teacher expectancy experiments, a teacher is given false reports about students' abilities, she might or might not change her behaviors toward the designated children depending upon how her own behavior pattern is expressed.

In our research, the effect has been largest for rural students and for suburban second graders, although for no case so far are these differences large enough to lead to statistically significant interaction effects. More work is needed to increase confidence that student subgroups are indeed as homogeneous as we have assumed in the present study. Among other things, there are rapid shifts in children's interests and capabilities over the grade-school

years. A task suitable for raising expectations at one level may be inappropriate, or relatively ineffective, at other levels.

Some implications of this work. As we noted earlier, the expectations which individuals have for their own performance can act as self-fulfilling prophecies; people who expect to do well often will do better than those who expect to do less well. In addition to this general effect, there are at least two more specific ways in which increased expectations may have important beneficial effects for children.

The first is that the behaviors which are associated with level of expectation—that is, the observable manifestations of an individual's expectation level often are also used by teachers to assess performance of children. Moreover they may be related to the actual quality of the child's activity in learning. For example, a child with high expectations is predicted by the theory to be likely to raise his hand often in class, to be more confident of his answers, and to be more willing to explore new or difficult subject areas. Since all of these behaviors typically are used by the teacher to assess performance of students, it seems reasonable to suppose that where other factors are equal, the higher the student's expectations for himself the more likely he is to receive favorable evaluations from the teacher, independent of any actual performance. Moreover, the behaviors associated with raised expectations are those associated with actual improvement of the learning situation. Children who are willing to enter new areas, who select difficult problems instead of easy ones, and who participate frequently and actively in class discussions probably actually do learn more than those who do not.

The second effect of raised expectations, a motivational one, may be a long-run increase in the child's general level of self-confidence about academic activities. This eventually could affect the sorts of college and career choices he will make. The higher the actual ability of other students in high school senior classes, the less likely is any given child to feel that he is performing in the top half of his class, independent of his actual ability or grades, (Meyer, 1970). Moreover, the independent effect of actual ability of others in determining likelihood of attending college in this same study was both negative and impressive: the gamma measure of partial association of others' ability and the college decision was —.11.

What Meyer reports is related to the relative expectation of students and suggests how the structure of competition may aid or hinder performance. Competition of some kinds may be a powerful incentive and lead to high levels of performance. Competition of other kinds may be debilitating. Competition can be employed rationally as an incentive for learning, but under pre-

sent conditions it is hard to know when or how. Being a member of a school class where other members have uniformly high ability probably lowers the individual's level of expectation for himself (independent of his actual ability), and the lowered expectation level for self in turn decreases the likelihood of attending college by a significant amount (again, independent of the actual ability of the child). Extending this theory, the higher a child's level of expectation for himself, the more likely he is to attend college, again, independent of his actual ability. If this proposition is true, then the effect of increasing children's expectations, by itself, should have measurable and beneficial results. However, it is impossible to know just how much treatment, and of what sort, would be necessary to produce the expectation changes indicated in this analysis.

A similar effect has been described by Davis (1966) as the "frog pond" effect. Among a group of college students of uniformly high ability (National Merit winners, finalists, and semifinalists), the higher the general ability level of the other college students, the less likely was the individual to choose to enter a "high-performance major field" (as opposed to vocational and "adjustment" majors). These results are consistent with those reported by Meyer, and it is reasonable to interpret them similarly in terms of expectation theory. The effect of having as significant others persons of uniformly high ability is likely to be the lowering of the individual's self-expectations and the consequent withdrawal from "tough" competition. People tend to avoid, when possible, activities in which they feel they will show up poorly. The businessman who is poor at golf meets his associates in the pool or clubhouse. By extension, if the expectations of a given individual could be raised, he would be more likely to enter the more challenging fields. Davis' results are especially interesting for, in view of his sample, it seems reasonable to conclude that any student had sufficient ability to enter a "high-performance" field; the actual differences in entry may be ascribed to differences in expectation for success.

At this time, we simply do not have sufficient knowledge from which to derive effective procedures for producing the generalized and enduring changes in expectations discussed here. However, we may note at least that these studies are consistent with predictions from expectation theory and that the consequences described are of major significance in the lives of the individuals studied.

References

Barber, J., and M. Silver.
1968 "Fact, fiction and the experimenter bias effect." Psychological Bulletin, 70:1–47.

Berger, J., and J. L. Snell.
1961 A stochastic theory for self-other expectations. Technical Report No. 1, Laboratory for Social Research, Stanford University, Stanford, California.

Berger, J., and T. L. Conner.
1966 Performance expectations, and behavior in small groups. Technical Report No. 18, Laboratory for Social Research, Stanford University, Stanford, California.
1969 Performance expectations and behavior in small groups. Technical Report No. 31, Laboratory for Social Research, Stanford University, Stanford, California.

Claiborn, W. L.
1969 "Expectancy effects in the classroom: a failure to replicate," Journal of Educational Psychology, 60:377–383.

Cohen, E.
1968 Interracial interaction disability. Unpublished research report available from the School of Education, Stanford University, Stanford, California.

Cohen, E. G., M. Lohmann, K. Hall, D. Lucero, and S. Roper.
1970 Expectation training I: Altering the effects of a racial status characteristic. Technical Report No. 2, School of Education, Stanford University, Stanford, California.

Davis, J.
1966 "The campus as a frog pond: an application of the theory of relative deprivation to career decisions of college men." American Journal of Sociology, 72:17–31.

Entwisle, D. R., and M. Webster, Jr.
1972 Raising children's performance expectations: A classroom demonstration. Social Science Research, 1:147–158.

Fleming, E. S., and R. G. Anttonen.
1970 Teacher expectancy or my fair lady. Unpublished paper presented at the annual meetings of the American Educational Research Association.

Jacobs, J. F.
1969 Teacher expectancies: Their effect upon peer acceptance. University of Florida, Gainesville, Florida, (Mimeo).

Kranz, P. L., W. A. Weber, and K. N. Fishell.
1970 The relationship between teacher perception of pupils and teacher behavior toward those pupils. Unpublished paper presented at the annual meetings of the American Educational Research Association.

Meichenbaum, D. H., K. S. Bowers, and R. R. Ross.
1969 "A behavioral analysis of teacher expectancy effect." Journal of Personality and Social Psychology, 13:306–316.

Meyer, W. J., and G. Thompson.
1956 "Sex differences in the distribution of teacher approval and disapproval among sixth grade children." Journal of Educational Psychology, 47:385–396.

Rosenthal, R., and L. Jacobson.
1968 Pygmalion in the Classroom. New York: Holt, Rinehart, and Winston.

Thorndike, R. L.
1969 "But you have to know how to tell time." American Educational Research Journal, 11:692.

Webster, M., Jr.
1970 Status characteristics and sources of expectations. Center for Social Organization of Schools, The Johns Hopkins University.

CHAPTER SEVEN

Raising Children's Expectations for Their Own Performance: A Classroom Application*

DORIS R. ENTWISLE

MURRAY WEBSTER, JR.

In this chapter we are concerned primarily with *consequences* of expectations: that is, with examining the behavioral effects produced by the expectation states actors already hold for each other. We are interested both in specifying more completely the types of effects that may result from these expectation states and in enumerating more of the empirical interpretations of

*Early work was supported partly by the Center for the Social Organization of Schools, The Johns Hopkins University, Grant No. OEG-2-7-061610-0207 and by Grant MH 18183-01 from the National Institute of Mental Health. Later work has been supported by the Office of Education, Grant OEG-3-71-0122. We would like to thank Margaret Boeckmann for her work devising the experimental procedure and Ellen Dickstein, Judy Kennedy, and Barbara Bricks for their help with the field work. Special thanks are due to Jack Epstein and Dr. George Gabriel for arranging for this research in schools. Principals who cooperated in the research helped in many ways. They are: Lyman Huff, William J. Maczis, Daniel Richowiak, Samuel Sharrow, Elliott Epstein, Ralph Thompson, Dr. Evart Cornell, Frank Tondrick, and George Fanshaw. We want to thank the teachers and students in the several schools who participated in the pilot studies and experiments.

specified consequences. To state the distinction somewhat differently, in preceding chapters one of the main interests has been specifying processes that would generate a particular pattern of expectations; in this chapter, a particular pattern of expectations is taken as the independent variable, and the interest is in specifying additional consequences, dependent variables, that will be produced by the expectations.

A second way in which the work reported in this chapter differs from that of previous chapters is that the interest here is primarily in *application* of the theory to empirical situations, rather than in theory development. Thus for this work we take propositions of Expectation States Theory that have received some verification in the highly controlled environment of the social psychological laboratory and attempt to use them to produce desired results in a naturalistic situation. This work may help reduce the gap between the precise but artificial laboratory setting, and the imprecise but naturalistic classroom setting in which educational research is often conducted.

Earlier chapters in this book have dealt primarily with problems of elaborating interactional and structural *determinants* of performance expectations; that is, with extending the basic propositions of expectation theory so as to predict the specific performance expectations formed from various combinations of initial status relations, agreement or disagreement interactions, unit evaluations of performances, or activating or making salient status characteristics. Most versions of the theory presented so far have included a statement of what has been called the "basic expectation assumption": the structure of expectations held by the group members will determine the distribution of the components of the observable power and prestige structure in the group. In chapter five, the source version of the theory was extended to include explication of the concept of a *source* of expectations, an actor whose evaluations are accepted and are used by others as the basis of the expectations they hold for their own and each others' performances. The basic expectation assumptions—Assumptions 3 and 4 in that chapter—were not problematic in this task, for they were assumed to have received adequate confirmation in earlier research.

In most (although not all) of the theory development research reported earlier the effect of expectation states produced by the various independent variables has been assessed using the single consequence *rejection of influence* under conditions of disagreement. For purposes of developing the theory by specifying additional determinants of expectations, this uniform measurement operation is an advantage, for it often permits direct comparisons of results across experiments. The basic expectation assumptions, however, actually predict a variety of behaviors (including rejection of influence) that result from expectation states, and these are important components of interaction as well. These other components of interaction are important both for explaining frequently reported results and for identifying variables that are important to

processes outside the immediate group participating in an experiment. For example, likelihood of acceptance of an action opportunity, a component of observable interaction that is predicted to vary according to the level of expectations held, is important not only as a component of interaction in a given group, but additionally because it is related to such other problems as learning in the classroom, assertiveness, and the impression of competence given to teachers and to other students.

Several investigators working with grade school children have reported results that in a general way are related to the variables of Expectation States Theory research. Rosenthal and Jacobson (1968), for example, report that when teachers are told that some randomly selected children were "potential academic bloomers," the selected children sometimes showed gains in academic achievement, both by teachers' ratings and by more objective measures such as intelligence test scores. Meichenbaum, Bowers, and Ross (1969) report similar results, and also show that teachers' behaviors toward the selected children differ from their behavior toward unselected children. By contrast, other investigators have reported clearly negative results from similar attempts to produce these effects (Claiborn, 1969; Jacobs, 1969). Using similar procedures in similar settings, they have been unable to produce similar effects. In this research, often called the "teacher expectancy" field, there are both successful instances and unsuccessful instances. This is perplexing. But even more perplexing from our viewpoint is that no matter what the outcome, there is no single, clear, convincing rationale to support the choice of techniques, the selection of the sample, the success of an experimental maneuver, or, in the cases where it applies, the failure to observe predicted effects.

The classroom research in expectancy concerns important phenomena, both in practical terms of children's learning and in theoretical terms of improving our understanding of educational institutions and the learning process, even though the findings have been sometimes inconsistent and contradictory. A major cause of the inconsistent results, we feel, is the lack of a sociological or social-psychological context. To our knowledge, work in the "teacher expectancy" field has not been guided in any systematic way by a theory that explicitly specifies the determinants, the definition, or the consequences of the "expectancies." We feel that Expectation States Theory, with appropriate interpretation, can specify the determinants, the definition, and the consequences of "expectancies," in this new research area.

The work reported in this chapter is intended to be a direct application of Expectation States Theory to problems of interest in educational research. Many ordinary classroom interaction situations meet the task orientation and collective orientation conditions of Expectation States Theory. Also, many of the "observable components of the power and prestige structure" specified in the basic expectation assumptions are similar to variables studied by educators.

Allocation of action opportunities, for example, can be seen when a teacher calls upon students.

I. THE TEACHER EXPECTANCY STUDIES; ANALYSIS AND INTERPRETATION

Studies in teacher expectancy research usually include the following features. First, the investigator describes his research to participating teachers as being involved with academic achievement of students and with predicting and assessing achievement. Students are tested, using one or more of the common standardized psychological tests. Test scores of students are not revealed to the teachers. At some later time, usually after a few days, the teacher is told that some students (typically around 20 percent of the class) have unusual academic potential, and that the teacher will probably observe unusual intellectual growth in those students during the coming months. The selected students are picked either at random, or in ways independent of their actual scores at the testing. Several months later, students are retested. In a recent reassessment of this research (Rosenthal and Rubin, 1971) about 39 percent of such studies show positive results: selected students show greater gains on retest than their unselected classmates (see, for example, Rosenthal and Jacobson, 1968; Meichenbaum et al., 1969). Positive findings have also been noted when the objective tests used to measure expectancy effects were administered by the schools rather than by experimenters.

Expectation States Theory suggests the following chain of events. The researcher enters a social institution where ability and evaluations of ability are central, and he presents himself as a capable judge of ability. The initial testing tends to legitimate his claims. The researcher, by stating that certain students possess unusually high academic potential, gives teachers information that should raise the expectations they hold for the chosen students.

If the teacher's expectations for a given student are raised, the basic expectation assumption predicts teacher behavior will be altered. The student will receive more action opportunities than other students, or than he received prior to the manipulation; for example, he will be called on more often in class. More importantly, this student will be more likely to receive positive evaluations from the teacher for any given performance output; that is, whatever the student says in class will be more apt to be positively evaluated. This student will be more likely to have his opinion agreed with by the teacher and by other students, especially in cases where the teacher and others have not yet decided upon the "correct" answer.

The effect upon a student's expectations of raising the teacher's expectations for him may also be predicted from the theory: his expectations for his own performance should be raised. The theory sees one of the immediate

determinants of any individual's expectations as the positive and negative unit evaluations of individual problem-solving attempts.

This analysis of the changes in teacher behavior points to selected students' receiving more positive evaluations, and as a consequence, coming to hold higher expectations for their own performance. Then according to the basic expectation assumption, the selected students will be more likely to emit performances. They will raise their hands more, and will speak out more in class. They will also be more likely to think their ideas are good ones, that their solutions to problems are the correct ones—in general they will be more self-confident and less likely to accept influence from others when their opinions are disagreed with. The process is circular, once set in motion, for these changes in students' behavior are of precisely the sort that are likely to lead teachers, even teachers who have forgotten the experimenter's initial revelations about test scores or other teachers who did not receive the score information, to convey high expectations.

Successful teacher expectancy studies are simply explained: students' higher grades on both standardized tests and classroom tests are the consequences of the expectation raising manipulation triggering the circular process described above. Higher scores on objective tests, for example, could stem from an improved mobilization of the student's resources in his now more responsive environment, from actual improved learning by greater class participation and involvement, or even from added increments of extra-school learning because improved self-confidence and positive expectations generalized outside school. First of all, learning should be improved by the increased interaction with the teacher and with other students, which would be directly predicted from the basic expectation assumption. Second, the behaviors associated with holding high self-expectations are probably important motivational amplifiers, perhaps leading to increased attention, curiosity, self-confidence, and interest in independent study.

The processes involved in the teacher expectancy research are important, both in practical terms and in terms of application of formal Expectation States Theory. We therefore designed several classroom studies in which the work to be reported was guided by two interrelated goals: (1) to apply some of the basic terms and assumptions of Expectation States Theory as a tool for analyzing a naturalistic situation, and (2) to use the theory to suggest simple procedures and tasks that will produce useful results in the field.

II. APPLICATION OF THEORY AND TASK DEVELOPMENT

For guiding the field studies in expectation raising, we decided to apply a version of Expectation States Theory that had previously received laboratory test and support. We are concerned here with ability and evaluations of

ability, expecially as they are affected by opinions of a "significant other" such as a teacher or a parent. Therefore, it seemed appropriate to adapt the single-source version of the source theory developed in chapter five. For reference, the explicit definitions and assumptions are presented in the appendix to this chapter.

All versions of Expectation States Theory assume the initial conditions of *task orientation* (interest in solving some problem) and *collective orientation* (willingness to consider answers from any individual in the group). These conditions are frequently met in classroom interaction: by definition, much learning activity is directed towards problem solution, and except for written testing situations, ideas and advice from many individuals are actively sought. The source theory speaks of a *source of evaluations*—an individual accepted as more competent to evaluate performances than the subject of interest, *p*. In the classroom, the teacher fills this role, both by virtue of his or her greater knowledge, and, usually, by access to objective information such as an answer key. For our experiments, we decided to fix acceptance of the experimenter as a source; we *told* children that we were competent to evaluate their performances at the task to be described below (information that they were willing to believe).

Given an accepted source, Assumptions 1 and 2 lead to the derivation that if the source (*e*) evaluates a series of performances by any individual (*p*), *p* will come to believe that his ability is consistent with the evaluations received. High ability conception (or self-expectation state) is the direct consequence of receiving positive evaluations from a source. Adding Assumption 4a enables us to predict an observable consequence of the expectation state *p* comes to hold; the higher the self-expectation state, the more likely is he to accept an action opportunity and to make a performance output.

Because we are interested in improving children's self-conceptions, we decided to attempt only to *raise* expectation states, not to lower them. (The latter would be desirable for theory testing purposes, not for our goal of practical application.) Thus, our experimental design calls for giving heavy doses of positive unit evaluations to selected children and predicting that this will raise their self-expectation states, and, consequently, increase their likelihoods of accepting action opportunities. In the classroom, teachers distribute action opportunities to the entire class when they ask something like "Who knows the answer to this question?" Children who think they know the answer—that is, children who hold high self-expectations for that task—accept the action opportunity by raising their hands or speaking out. Our experiment, described in detail below, was designed to be analogous to this sequence of behaviors. In general, we predict that we can increase the rate at which a child raises his hand in response to group-directed questions as a direct function of the self-expectations that he holds, and that we can increase the child's self-expectation state by giving him a large number of positive unit evaluations

of past performances. More precisely, we formulate for testing the following two derivations, using Assumptions 1, 2, and 4a from the single source theory:

Derivation 1

If an individual (p) has received no unit evaluations of his performances at time t_1 and receives a large proportion of positive unit evaluations from a source at time t_2, then the likelihood that p will accept a given action opportunity and make a performance output will be greater at time t_3 than at t_1.

Derivation 2

If an individual (p_1) has received a large proportion of positive evaluations at time t_2, then as compared to a second individual (p_2) who has not received any performance evaluations, the likelihood at t_3 that p_1 will accept a given action opportunity and make a performance output is greater than the likelihood that p_2 will do so.

A large part of our initial work was devoted to development of an appropriate experimental task that had to meet different requirements from those met by previous laboratory tasks. Laboratory studies usually alter expectations by giving subjects false information. For example, subjects are given fictitious test scores or are led to believe that other subjects are disagreeing with them by means of apparatus that alters communications. All previous laboratory work and all teacher-expectancy studies mentioned earlier rely heavily upon the use of false information to alter expectations.

Using false information may be very effective and consequently useful for research purposes (especially in laboratory studies where the deception involved may be explained immediately afterward), but for repeated use in applied research false information is clearly not desirable. In studies where expectation states are raised, there are both practical and moral difficulties associated with the continued use of false information. For example, one would not wish to tell a child that he is bound to do much better than he has been because if the prediction fails he may suffer an impaired self-image. As for giving teachers false information, the Rosenthal and Jacobson studies have already received sufficient notice in the press so that teachers and principals nod smilingly if an investigation reports "new high test scores." In studies where expectation states are relatively lowered, as in some tracking studies, there are moral difficulties. No one would care to tell a child that he is likely to fail in the next semester or that he will not do as well as he hopes. Research in a naturalistic setting thus demands modification of the main experimental maneuver used in the previous laboratory work with the need to alter expectation states in some way *other* than the giving of false information.

After some pilot studies we decided upon the procedure described in what

follows, a modification of a story-telling task previously used in research on cognitive development of children (Entwisle, Grafstein, Kervin, and Rivkin, 1970). One adult interacting with one child gives consistent positive evaluations of performance in a task where the child's actual ability is almost irrelevant. The evaluations are therefore not inconsistent with anything known about the child or with his potential. The aim of the maneuver is to raise the child's expectations. In addition to meeting the major criteria listed above, the pilot studies demonstrated that the task met practical criteria: (1) it provided discrete, easily observable performance outputs that permit clear evaluations, and (2) it was interesting enough to capture children's attention.

The experiment has three phases, corresponding to the three "times" referred to in the derivations. Phase I determines the "baseline" level of acceptance of action opportunities for both the "experimental" and the "control" groups. Children fill in words in a story skeleton and the experimenter is neutral. The experimenter notes who raises his hand as each word is called for. (The experimental and control groups correspond to individuals p_1 and p_2, respectively, in Derivation 2.) In Phase II, the attempt is made to manipulate upwards the self-expectations of children in the experimental group. One child (experimental group child) fills in words in a second story skeleton and is praised and encouraged by the experimenter after every word. The other children (control group) participate in a neutral procedure. Then the group of children (experimental and controls) is reassembled for Phase III and fills in a third story skeleton, with the experimenter again neutral and noting who raises hands.

To tie this back to the theory: the validity of Derivation 1 can be assessed by comparing the results of Phase I to Phase III for children in the experimental group. How well Derivation 2 is confirmed can be assessed by comparing results of Phase III for children in the experimental group with results of Phase III for children in the control group.

The basic experiment just described serves as the foundation for all empirical studies to be reported in this chapter, although modifications are introduced where necessary for purposes of studying the effects of new variables. Experiments have been classified into three "series," according to chronology and according to issues that emerged during the course of the research.

For the First Experimental Series, work is concerned with testing the task and experiment designed: would they work in the way we expected on the basis of the theory and laboratory studies? Would the children understand and be willing to participate in the research? Is the procedure simple enough that we could show others who have little interest in abstract theory or experimental design how to raise expectations? In order to gain information on these and other basic questions, we applied the procedure to as wide a range of subjects (white rural children; black inner city children) as initially pos-

sible, in the age range of some interest in expectation development (third and fourth grades).

The Second Experimental Series was conducted to expand downwards the age range to study earlier effects upon expectation state development, and white middle class children were the subject sample. This extension permits comparison with data from the First Series to assess the relative effects of status, race, age, and sex factors upon the processes.

The Third Experimental Series is addressed to "special problem" issues that arose in the earlier work: the effect of racial mismatch between source and child, the "debilitating effect of school" in forming low self-expectation states, the "contrast effect" in our expectation raising experiment, and an alternative "behavioral modification" interpretation of our experimental results.

III. FIRST EXPERIMENTAL SERIES

A. SUBJECTS

For the first set of studies, children were drawn from four schools in the Baltimore area, two inner city schools with nearly 100 percent black students, and two rural schools with nearly 100 percent white students. The rural schools are located in a farming district 30 miles north of Baltimore, near the Pennsylvania line.

For the two inner city schools, only students whose school records showed tested IQ scores between 90 and 110 were selected for the study. The IQ range for rural students is considerably larger, from 76 to 141, but the mean IQ is 105 for rural third graders and 108 for rural fourth graders. Experimenters were middle class white persons in the 20–30 age range, both sexes.

B. PROCEDURE

At the beginning of an experimental session, children were brought together and told that the researchers were looking for people who could tell good stories. They were to be divided into "teams" and were told that the team that made the best stories would win a prize. Then one experimenter took the members of each team to a separate room and described the story-telling task to them. Members of a team were chosen so that children on one team came from different classrooms.

Children were told that the "game" consisted of making up a story. The experimenter would help by starting sentences, but then the children should

try to make interesting stories by supplying "good" words when asked for words for the story. In every phase the same story skeleton (see Figure 7-1) was used, and the skeleton contained twelve blanks. Children filled these blanks with twelve words or phrases.

THERE WAS ONCE A VERY TALL PRINCE WHO HAD A (castle)
THAT (HE, SHE) (lived in)
ONE DAY (HE, SHE) HAD TO GO TO (the dungeon to see his prisoners)
(HE, SHE) DID THIS VERY (angrily)
BECAUSE (HE, SHE) WANTED TO (make sure they were there)
THIS WAS VERY DANGEROUS BECAUSE OF THE (strong prisoners)
WHO (WHICH) WAS (WERE) VERY (mean)
IN ORDER TO FOOL THE (FILL IN) THE (FILL IN) DRESSED UP AS (another prisoner)
IN SUCH A DISGUISE THE (FILL IN) LOOKED (mean)
AND WHEN THE (FILL IN) SAW THE (FILL IN), THEY (welcomed him)
THIS MADE THE (FILL IN) (feel pretty good)
AND /COMPLETE STORY/ (he let his new friends go).

The experimenter chooses from alternatives in parentheses the item consistent with the story line. For example, in this story the pronoun "he" is chosen because it refers to "prince."

FIGURE 7-1
Story Skeleton with Sample Entries from a Rural Group

Children were told to listen carefully while the sentence was being read, then when the blank was reached, to try to think of a good word. Anyone who thought of a good word was to raise his hand, and the experimenter would select one child to give the "team's word" for that sentence. Children were cautioned not to raise their hands unless they thought they had a good word, for if they were called on and gave a bad word, this would hurt the team's score. The purpose of this instruction was to help maintain the task orientation and collective orientation required by the scope conditions of the theory. The experimenter allowed 30 seconds to elapse after reading a sentence before calling upon a child.

Before calling on a child, the experimenter recorded privately which children were holding up their hands (the measure used to determine the expectation state—acceptance of an action opportunity). The experimenter held a clipboard so that children could not see what was on it. The clipboard was used for recording words given by the children. During the 30-second waiting period, he made small marks indicating which children had volunteered without the students being aware of this.

Phase I consists of an initial story being produced as just described. The experimenter (E) does not evaluate any of the words given during Phase I. He calls upon each child in Phase I approximately the same number of times. With twelve blanks to fill in and with four children playing, each child can be called upon three times. In only a few instances did any child volunteer fewer than three times altogether and so unbalance the selection of respondents.

At the end of Phase I children are asked to return to the room where they initially assembled. After they begin to move out, E quietly tells one child (chosen because his level of hand-raising was near the median for the group) to stay in the experimental room and wait there a minute. After E makes sure that the control children are on their way to the proper destination, he returns to the room and to the selected (experimental group) child. E then tells the experimental child that he will have an opportunity to make up a story all by himself; also that E has played with many children making up stories and that E thinks he/she (the experimental child) is really good at the task. Then a story skeleton with a new lead word is filled in orally by the child, just as before, but of course with only one child producing the story there is *no* volunteering. The child merely supplies a word for the blank when the sentence is read by E. After each blank is filled, E indicates approval vigorously —by smiling, by nodding, by commenting "a very good word"—"good"— "that's interesting!", etc., that is, he indicates approval in every possible way consistent with sincerity.

Several experimental groups (four or five) were run simultaneously with an E for each group. A single coordinator for all the groups managed the initial explanation, story-reading for the Phase II control groups, and prize award sessions. When Phase II for the experimental child was complete, he went and joined the control children in their Phase II. Phase II for control group children is a story-reading session, and control children from several game groups gather in a central room as they finish Phase I to listen to the coordinator read a story. Children from the several experimental groups join the story-reading group (control groups, Phase II) as they finish their Phase II activities. The experimental children thus listen to the end of the story being read to the control groups. The experimental children's entrance is not noticeable because they join the story-reading group while the children's attention is directed toward the story teller. The story-reading prevents communication among the children during Phase II. At the end of the story-reading the children are asked to "go back to the room where you were before." *All* children thus return together. At this point E's are rotated among rooms so each E has a new group and is unaware of the identity of the experimental child. Phase III consists of a repetition of Phase I with the experimenter noting how many times each child volunteers to supply a word. The length of each phase varies, as would be expected, but the Phase II control procedure can adjust its length to the time requirements of each set of experiments.

At the conclusion of Phase III, the children are again brought together in a large group. One team is selected at random as having constructed the best team stories, and each member of the winning team is given a prize of a regular size candy bar. All other children are given a miniature candy bar, thanked for their help, and escorted back to their classrooms. Figure 7-2 summarizes the experimental design for all this research.

Results

Table 7-1 presents the gain in the number of times children in the control groups and in the experimental groups volunteered by raising their hands from Phase I to Phase III. (The mean *number* of times children raise hands in all groups is about 6.7 in Phase I.) A t-test of the mean gain of the experimental groups vs. the mean gain of the control groups is highly significant ($P(t_{270} \geq 3.17) < 0.01$). Other t-values are displayed in the table for various subdivisions of the sample. It is noteworthy that in each stratum the gain for the experimental group exceeds the gain for the control group, even though not every individual comparison is statistically significant on t-test.

TABLE 7-1. Gains in Rate of Volunteering from Phase I to Phase III. (N's in parentheses)

	Experimental Subjects		Control Subjects		Probability level, One-sided t-test of Differences between Experimental and Control Groups
All Black	(36)	+1.22	(108)	+0.64	N.S.
3rd Grade	(20)	+1.85	(60)	+0.88	N.S.
4th Grade	(16)	+0.44	(48)	+0.33	N.S.
All White	(32)	+2.28	(96)	+0.36	0.001
3rd Grade	(22)	+2.45	(66)	+0.56	0.001
4th Grade	(10)	+1.90	(30)	−0.07	0.025
Grand Total	(68)	+1.72	(204)	+0.51	0.001

The success of the story-telling task may also be assessed by noting the proportion of children in the experimental groups whose rate of hand-raising increased from Phase I to Phase III compared to the proportion in the control group (Table 7-2).

The t-tests assess the magnitude of the gain but this increase could come about because some children's rates increased markedly even though others did not. By examining the proportion of children who gain, one has information on the consistency of the effect. Proportions i the experimental groups consistently exceed proportions in the control groups. The overall proportion

TABLE 7-2. Percentage of Persons Showing Gains from Phase I to Phase III.

	Experimental Subjects	Control Subjects
All Black	68	53
3rd Grade	78	68
4th Grade	56	33
All White	86	55
3rd Grade	89	61
4th Grade	80	43
Grand Total	76	54

of children gaining in the experimental groups is 76 percent, compared to 54 percent, just about chance level, in the control groups.

Thus the predictions of Derivations 1 and 2 are in general borne out by the results of the research. The data of Tables 7-1 and 7-2 show that the procedure apparently produced increases in expectation states, as measured by acceptance of action opportunities, although success was much greater with [illegible]en than with black children.

C. ASSESSMENT OF FIRST EXPERIMENTAL SERIES

In the First Experimental Series, 76 percent of the children who received the experimental treatment showed the desired increase, a significant gain when compared with that made by control students. Operationally the experimental procedure was a success. From a practical standpoint, the procedures were simple and easy to use. Field notes indicate that the children found the task involving and enjoyable and were highly motivated to succeed at it.

Variability in gains resulting from using this experimental procedure deserves some comment. (The black children have already been commented upon.) As a child grows older, his expectations for his general performance level at most tasks probably crystallize. Thus a greater increase for third graders than for fourth graders in both the control and experimental groups is not surprising. In fact such considerations led us to select third and fourth graders initially rather than older children who might have been preferable as subjects on other grounds. It also led to our extending the age range downward as reported in the next section (Second Experimental Series).

As mentioned earlier, maneuvers that give fictitious results to students or to teachers have drawbacks. Such maneuvers may be defended when they are one-shot procedures to demonstrate self-fulfilling prophecies, placebo effects, and the like, or when they are needed to allow quick and sizeable manipulation of variables in the laboratory. But such maneuvers are not defensible over the long term. They are not even viable when the goal is to change children's long-range expectations for themselves or others' expectations for

children in classroom settings. If one goal of education is expectation alteration, then ways of altering expectations must be found that are compatible with other educational goals and that are suitable over the long term. The procedure described here has achieved modest success along these lines.

Besides the practical and educational criteria considered above, one must also evaluate the procedure in terms of its ability to operationalize the variables of the theory. Probably both "experimenter effects" and unintended biasing arise. Third, and perhaps most important, the question of alternative interpretations may be raised. Does Expectation States Theory provide the best context for interpreting the present results? We will now analyze these three questions in turn, and in the section on "Special Control" experiments we will reconsider them in the light of additional data.

(1) *Experimenter effects* were equalized insofar as possible. First, E's were trained to treat the children as equally as possible in Phase I and to refrain from evaluating any performances. The child selected as the "experimental" child at the end of Phase I was chosen on the basis of his having responded close to the median for his group. To choose a child with too *low* a response rate in Phase I might have biased results in favor of predictions through a "regression to the mean" phenomenon in Phase III. To choose a child with too *high* a response rate in Phase I would have biased results against predictions, because of a "ceiling effect."

To avoid drawing attention to the child selected to participate in Phase II, E did not say anything to that child until the children were on their way to the other room for Phase II; nor did the E explain at all to the others why this child was asked to remain behind. In most cases the fact that one child remained behind was apparently not noticed.

In addition, the control children were occupied at similar tasks in Phase II. *All* children were with an experimenter during Phase II, because control children were with the research coordinator in a story-reading session. The experimental group children participated in at least part of the control treatment because they joined the story-reading group as they finished the Phase II experimental treatment.

(2) *The issue of biasing* is troublesome, for even a slight change in a teacher's manner will change a child's disposition to raise his hand. Therefore, a number of steps were regularly taken to minimize bias from E. The major danger of biasing would come in Phase III, however, when E might respond more warmly or more positively to the "experimental child" than to the others. This child might also feel that because of the individual session with E in Phase II he (the child) had some "special" relationship with E. In order to minimize problems of this nature, E's were rotated before the beginning of Phase III. Thus the E was new to the entire group in Phase III, unfamiliar to both the experimental and control children. Furthermore, the *E did not know at this point which child had been given the Phase II "treatment"* and thus was

unlikely to treat the children differentially. In Phase III, as in Phase I, E's are trained to call on every child the same number of times insofar as this is possible.

(3) *Alternative interpretations* of experimental results are relatively easy to invent. Certainly we cannot rule out all competing interpretations for the results reported here. It seems difficult, however, to contrive an explanation that accounts for *all* our results, as well as for work by others mentioned earlier, as satisfactorily as Expectation States Theory. Suppose, for example, the effect of the treatment of the experimental group child in Phase II was to offer additional practice in story telling, and that practice *per se* increased the child's confidence in his ability to tell stories. This would imply that the strongly positive Phase II evaluations were unimportant, and possibly that the level of expectations in Phase III was irrelevant to behavior (hand raising) as well. But the task was selected partly because it calls for no special ability; it certainly is not one that can be "improved" through practice. Without evaluations from E, children have trouble deciding whether their words are "good ones": children's remarks in the "unevaluated" Phase I point to this. So practice as an explanation is not appealing.

The results, by way of another example, might be accounted for by seeing the Phase II procedure as *reinforcement of behavior* rather than *positive evaluation*. The experiment might be seen as demonstrating that children will increase the rate of emission of behaviors that have been rewarded in the past. In assessing the reinforcement explanation, the reader should note carefully that our measure of expectation states in Phase I and Phase III was rate of volunteering, or *hand raising*. What were positively evaluated in Phase II were *words spoken* by the children. At no time during the experiment was anyone positively evaluated (or reinforced) for hand raising, the measured behavior. Positive evaluation (or reinforcement) was *never* given for hand raising because during the Phase II experimental treatment the child did not raise his hand. In the other phases when hand raising was occurring no evaluation (or reinforcement) was given. Furthermore, the specific words evaluated in Phase II were seldom the words volunteered in Phase III. Thus, a reinforcement explanation lacks force on analytic grounds; it requires making several tenuous interpretations of experimental variables, and making some rather complex assumptions about stimulus and response generalization.

Other research that explicitly measures children's expectations by using a kind of self-rating scale is consistent with our results. Expectations are seen to increase following approval and positive evaluation by an adult, but are unchanged when an adult maintains a neutral role (Hill and Dusek, 1969; Crandall, 1963; Crandall, Good and Crandall, 1964). The actual procedure used by Hill and Dusek was very similar to our own, for an adult responded "That's good. Fine. Very good. You're doing well" for positive evaluations following attempts at an angle-matching task. The adult was neutral and

nonresponsive for the nonevaluative condition (like our Phase I and Phase III procedure). Other studies, also consistent with the conclusion, show that if an individual gets approving reactions from others with respect to some specified attribute, he will improve his self-rating on that attribute (Maehr, Mensing, and Nafzger, 1962; Videbeck, 1960).

IV. SECOND EXPERIMENTAL SERIES

The overview of the first set of classroom studies is encouraging, both in terms of procedure development and theory application. The Second Experimental Series aimed to extend the age range and dealt with middle class children. We wished to determine the age limits within which experiments would be robust and wished to see whether sex and race subgroups responded differentially to the procedure (hinted at by the apparently lower susceptibility of black children). Also, of course, further experiments provide replication of earlier work.

A. SUBJECTS

The experimental procedure already described was used with a sample of children from a white middle class suburb of Baltimore. There were approximately equal numbers of boys and girls, with 79 first graders, 84 second graders, 112 third graders, and 103 fourth graders. Insofar as possible, the four members of each experimental group were chosen from different classrooms. Some of the grades were distributed among four classrooms, some among only three. In the latter case, two children were taken from a single classroom, and the remaining two from two other classrooms.

B. RESULTS

For this and later experiments, we now change the method of reporting results and report results of analyses of variance. With a factorial design where treatment (experimental vs. control), grade, and sometimes residential locus are considered to be three fixed-effect factors, the data can be revised to yield proportional subclass numbers by randomly discarding some experimental groups. One second-grade group, eight third-grade groups and six fourth-grade groups were therefore eliminated using a random process before analyzing the data shown in Table 7-3.

TABLE 7-3. Average Gain in Rate of Volunteering From Phase I to Phase III, White Middle Class Students (*N*'s in parentheses)

	1	2	3	4	Av.
Experimental Group	1.20	2.55	1.85	1.85	1.86
	(20)	(20)	(20)	(20)	(80)
Control Group	0.67*	0.67	1.20	1.38*	0.98
	(59)	(60)	(60)	(59)	(238)
Difference Between Groups	0.53	1.88	0.65	0.37	0.88.

*This mean is based on 59 rather than 60 observations.

Table 7-3 shows the mean gain in rate of volunteering from Phase I to Phase III by grade and sex for treated (experimental group) and untreated (control group) middle class children. Groups are approximately balanced for sex. Three sorts of changes in the dependent variable are reflected in this table. First, all experimental and control groups show some increase in rate of volunteering. Second, for all grades the increase is greater for children in the experimental group than for those in the control group. Third, second graders show the greatest difference between the experimental and control conditions. These results all are consistent with results of the First Experimental Series.

The analysis of variance in Table 7-4 based on gains in rate of volunteering from Phase I to Phase III reveals a significant treatment effect ($p < 0.05$), no significant grade differences, and no significant grade x treatment interaction.

TABLE 7-4. Analysis of Variance for Gain in Rate of Volunteering, White Middle Class Students. (Approximately balanced for sex)

Source	d.f.	Sum of Squares	Mean Square	*F*-value
Experimental vs. Control Treatment	1	46.81	46.81	6.02*
Grade	3	22.42	7.47	0.96
Treatment × Grade	3	20.27	6.76	0.87
Within Treatment × Grade Groups	310†	2409.70	7.77	

*Beyond the 5 percent level
†Two observations are missing, estimated by subclass means.

It is of interest to combine data from third- and fourth-grade white middle class subjects of this experiment with the rural and inner city data discussed in the First Experimental Series. To do this, the earlier data have also been reduced by randomly discarding six black inner city fourth-grade groups and two white rural third-grade groups (see the data summarized in Table 7-2 vs. data of Table 7-6). Some of the fourth-grade subjects included in Tables 7-3 and 7-4 have also been eliminated randomly.

For the combined subjects of the First and Second Experimental Series, the treatment effect is again judged highly significant ($p < 0.01$). There is no evidence here that the effect of the treatment is different by grade or residential locus, since none of the interactions of treatment with any other factor is significant.

TABLE 7-5. Average Gain in Rate of Volunteering From Phase I to Phase III, for Children from Three Residential Loci, Grades 3 and 4 (*N*'s in parentheses)

Residential Locus	Suburban		Inner City		Rural	
Grade	3	4	3	4	3	4
Experimental Group	1.85	2.20	1.85	1.90	2.60	1.90
	(20)	(10)	(20)	(10)	(20)	(10)
Control Group	1.20	2.07	0.97	0.20	0.57	−0.07
	(60)	(30)	(60)	(30)	(60)	(30)
Difference Between Groups	0.65	0.13	0.88	1.70	2.03	1.97

The combined data show that expectations were raised significantly in experimental children compared to control children, and as shown in the other analyses (Entwisle and Webster, 1973a), residential locus probably does not affect susceptibility to treatment (no $T \times R$ interaction).

TABLE 7-6. Analysis of Variance for Gain in Rate of Volunteering; Three Residential Loci; Two Grades

Source	d.f.	Sum of Squares	Mean Square	*F*-value
Treatment (T)	1	110.21	110.21	14.75*
Residential Locus (R)	2	43.62	21.81	2.92
Grade (G)	1	0.87	0.87	0.12
$T \times R$	2	22.21	11.10	1.49
$T \times G$	1	0.50	0.50	0.07
$R \times G$	2	35.77	17.88	2.39
$T \times R \times G$	2	2.27	1.14	0.15
Within T, R, G Subgroups	347†	2593.15	7.47	

*Beyond the 1 percent level
†Two observations are missing; estimated by subclass means.

C. DISCUSSION

The results of the Second Experimental Series extend the positive findings of the First, since white middle class children also respond to the experimental procedure by increasing their rate of hand raising. Although there are other groups that could be studied (for instance black middle class), the range of children already included suggests that the phenomenon is replicable and fairly general. The parallels are clear between the experimental task, comple-

tion of a story under the experimenter's direction, and many tasks overseen by teachers in classrooms. Expectation States Theory has, then, at least some relevance to elementary education.

The significance of social status in this study differs from its significance in earlier Expectation States Theory work. Generally, previous status characteristics research has been concerned with the effect of *differential status in assignment of power and prestige within a small group*. E. G Cohen (1968, 1970), for example, has studied mixed groups of black and white boys attempting to solve a problem together, to see who makes more performance outputs. In the laboratory, Webster (1970) has studied the effect of status characteristics on the effectiveness of evaluators; performances were monitored by evaluators of high status (college students) or of low status (eighth graders). In both E. G. Cohen's and Webster's studies, it has been shown that the status characteristic will, under certain circumstances, lead to differential conceptions of ability in accord with the different states of the diffuse status characteristics.

The focus of the present work differs. Here we have groups, all of whose members are *equal* with respect to a status characteristic (such as sex, age, and race). The question is whether expectations will be raised to the same extent for members of each status group. Can expectations of rural children be increased as easily as expectations of inner city children, for instance? The intent is thus to look at the "demography" of expectation raising, particularly for the kinds of children—rural or inner city—who are often classed as low achievers. This experimental treatment for modifying children's expectations has worked with children from three very different residential settings and no significant differences by residential locus have appeared. The effect has been greatest for rural students and the suburban second graders, but not significantly greater (no significant treatment × locus interaction). So far age does not appear to be important in expectation raising, although only for one group (middle class) has age been sampled over any sizeable range.

More work is needed, however, to increase confidence that groups are as homogeneous as suggested here. For one thing, over the grade school years there are rapid shifts in children's interests and capabilities, so that a task suitable at one level may be inappropriate, or relatively ineffective, at other levels. The present research assumes that the group is collectively oriented and seriously motivated towards high performance of the task at hand. Partial failure to meet these conditions would attenuate any observable effects of expectations, or of our attempted experimental manipulations.

V. THIRD EXPERIMENTAL SERIES

The focus of concern in the First Experimental Series was task development and general feasibility of the research program. The Second Experimental

Series focused upon determining the generality of the effect across societal status groups. In the Third Series we were concerned with pursuing some suggestions, and also possibly problems, that grew out of the earlier experiments. Some instances of possible refractoriness to the expectation-raising treatment had occurred and these deserved more study. In addition, alternative interpretations of the results could be investigated with further work.

In studies reported to this point two kinds of children are not very responsive to the treatment: inner city black children and white middle class girls of third and fourth grades. The results for inner city children have already been discussed at some length. The lack of response in middle class girls has not been obvious to the reader because it was not possible to include sex as a factor in the analysis. The reader will note, however, in Table 7-7 where the full set of data for third- and fourth-grade suburban children is presented and tabulated by sex, that boys show sizeable increases and girls actually show decreases.

TABLE 7-7. Average Gain in Rate of Volunteering From Phase I to Phase III for White Suburban Children, Third and Fourth Grade. (*N*'s in parentheses)

Girls			Boys		
Experimental	Control	E-C	Experimental	Control	E-C
		Grade 3			
0.36	0.86	−0.50	2.29	0.71	1.58
(14)	(42)		(14)	(42)	
		Grade 4			
1.08	1.46	−0.38	1.86	0.52	1.34
(12)	(35)		(14)	(42)	

Also, during the experiments the experimenters observed that the white middle class girls were particularly withdrawn and unresponsive. We therefore performed two replications.

1. In the first replication we returned, exactly one year later, to the inner city schools where we had conducted the first experimental series with black children. On the return visit we employed only black female experimenters. Everything else in the replication including time in the school year, was the same as in the First Series.
2. The second replication study consisted of experiments with only girls of the third and fourth grades in a white middle class suburban school. The school where the replication was carried out was different from the school in the Second Series, but was very much like the Second Series school in social class and other characteristics. (Further work could not be done in the first school because all students had already participated.)

A. BLACK INNER CITY REPLICATION

Two general lines of thinking seem to be consistent with the relatively lower efficacy of the experimental treatment for inner city black students. The first, which we may call the "debilitating effect of school hypothesis," holds that an important outcome of school for black children is to lower their self-confidence and self-evaluation. We speculate that as black children go through the grades, they receive predominantly negative evaluations of performances, and consequently become progressively more certain that they will fail at anything they attempt. They come, in other words, to hold self-expectation states that are fixed and low. If such effects have been building up over several years in school, our experimental procedure may be just too weak or too short in duration to produce any marked expectation raising with these children. The fact that effects of the procedure vary inversely with school grade for all three status group samples in the First and Second Experimental Series is consistent with the idea that expectation states become more difficult to modify as children become older.

An entirely different explanation is suggested by the work of Katz and his associates (1968, 1970), who present indications that white adults are perceived as hostile by some black children, and that when white examiners test black children, the children often assume that they are being compared to white children. In the former case, our white experimenters may not have been accepted as "sources" by the children; that is, as "significant others," whose evaluative opinions are accepted by them. In the latter case, we would expect that children would form low self-expectations because of the inferred comparision to white children who possess higher diffuse status, in the manner described theoretically by Berger et al. (1966), and documented with grade school children by E. G. Cohen and her associates (1970).

One way to distinguish between these competing explanations is to repeat the earlier research using black experimenters. If the "debilitating effect of school" is the explanation, then black experimenters should also be ineffective in raising expectations of black children. If racial mismatch was the problem, then black experimenters should be effective.

For this series, black women, students at a nearby college, served as the E's. As already mentioned, the schools selected for the experiments were the same inner city schools as those used the previous year. Students were all black fourth graders, the most resistant group in the earlier study, and also the group best suited for distinguishing between the competing explanations. Of course they were not the same students as those who had previously participated.

Table 7-8 shows the gain scores for children using this sample and the black experimenters. Control group children show an average gain of 1.82, and the

average gain in the experimental group is 3.37. A *t*-test of the difference in gains is significant beyond the 5 percent level.

TABLE 7-8. Average Gains in Volunteering; Black Ss; Black Es

Group	*n*	Gain
Experimental	57	+3.27
Control	19	+1.82

Black experimenters were thus successful in producing a gain in the rate of volunteering of black children. In previous data for inner city blacks with white experimenters, gains were not significant (see Table 7-1).

A number of reasons could account for the discrepant results of the two experiments with black inner city children. The most appealing reason is the racial difference in experimenters from the first experiment to the second, but a much more tightly controlled set of experiments is required to rule out other explanations. The experimental procedures, for example, were considerably refined by the time of the repeat experiments. Also, the black E's were probably more alike and more uniformly skillful than the E's used in the First Experimental Series. Nonetheless, significant differential gains are seen in a subgroup that previously had not manifested significant gains, and this argues once more for the general effectiveness of the experimental procedure.

Temporarily being less cautious, we note some suggestions based upon these results and the ideas mentioned earlier from Katz's work. This work indicates that some E's may not possess the necessary attributes to serve as effective sources. Attributes of a source, besides higher status (white skin color and/or age) or higher ability, reflect on that source's effectiveness as a purveyor of expectations, no doubt. Perhaps some sources can never achieve complete credibility for some subjects even though the status or ability criteria are met. An example will make this clearer. Suppose two social scientists are colleagues, and one is highly eminent and nationally known, whereas the other has only a local reputation. The eminent colleague is disposed never to criticize and always makes favorable comments about the work of his local colleagues. In this situation the less widely known social scientist will not weigh very heavily any favorable remarks from his eminent colleague because he sees very little variation in the tenor of the eminent colleague's remarks; no matter what occasions them, they are always positive. The eminent colleague does not possess credibility in the sense we have used the term even though his status and other attributes, as well as his positive remarks, meet the surface conditions of serving as a source. In the same vein, if black children interact mostly with adults (white teachers who elect to work in the ghetto) whose strong tendency may be to remain pleasant and to refrain generally from negative evaluations, such adults may lose credibility as sources for raising expectations.

The issues here are undoubtedly complex. Studies by Katz and his associates (see Katz, 1970) suggest that a black examinee's perceived probability of a successful performance on a test is apparently determined by his beliefs about the reference group he is being compared with—if black comparison groups are explicitly mentioned, the black examinee does best with a white administrator. In our use of white E's with black S's in the First Experimental Series, the children might have assumed they were being compared with other children (mostly white) whom the white E had dealt with in the past. In this case, following the line of reasoning suggested by Katz's work, the probability of success perceived by black children may have been low.

In applied work the characteristics of a source that makes him effective at raising expectations may be a research issue of high priority. Teachers, for example, of the same ability (educational level) and status are notoriously variable in their influence on students. One thing making a teacher effective may be ability of the teacher to hold and to convey high expectations about students. A teacher's high expectations for a student may be a powerful inducement for the student to enter a high expectation state for his own performance.

B. SUBURBAN FEMALE REPLICATION

At the beginning of this section (see Table 7-7) it was noted that third- and fourth-grade suburban girls actually decrease their rate of volunteering from Phase I to Phase III. When these data are pooled with other suburban data and sex is not included as a factor in the design, increases in other groups are sufficient to mask this finding. Overall there is a significantly larger gain score for experimental subjects versus control subjects in spite of the negative gains observed for these girls. Since research in an early stage, like the exploratory studies presented in this chapter, cannot be entirely insensitive to patterns in the data, we obtained permission to run an additional fifteen experiments ($n = 60$) with third and fourth grade *girls only* at a suburban school much like the one enrolling the suburban children of the Second Experimental Series. These results are shown in Table 7-9.

TABLE 7-9. Average Gains in Volunteering: White Suburban Female Subjects; Grades 3 and 4

Group	n	Gain
Experimental	15	+1.80
Control	45	+1.49

For experimental group girls in this replicating experiment, the mean gain in rate of volunteering from Phase I to Phase III is 1.80; for control group

girls it is 1.49. The difference, 0.31, is not significant on *t*-test. Our tentative conclusion is, therefore, that the present procedure has not been effective for third and fourth grade white suburban girls, although we now believe it is probably effective for all other groups, if races of experimenters and subjects are matched.

What factors could account for the sex difference in effectiveness? It is thought that boys, more often than girls, are blamed or criticized by teachers of elementary school (Brophy and Good, 1969, 1970). Because of this, the expectation-raising procedure of Phase II may be considerably more effective for boys than for girls. Boys who hear consistent praise during the experiment may be contrasting it with a background before the experiment of little positive evaluation. Girls, on the other hand, who apparently are seldom subjected to blame or criticism, could interpret the expectation raising maneuver of Phase II as a continuation of positive evaluations they have been receiving all along. Or, since girls generally do well in elementary school compared to boys, their expectations for performance at verbal tasks may already be at an asymptote so that attempts to raise expectations further are fruitless. Our data do not allow a choice between these explanations.

C. SPECIAL CONTROL EXPERIMENTS

In assessing the basic experimental procedure tested in the First Experimental Series, we concluded that the Expectation States Theory interpretation of the results was more plausible than alternative interpretations. The most reasonable alternative interpretation seems to be one that asserts that the experimental procedure produces increases in hand raising, not because of any improvement of the child's expectation state, but rather because of rewarding features of the Phase II situation. The alternative interpretation consists of one or both of two basic arguments: (1) experimental group children are responding to the situational rewards of receiving extra attention from an adult during Phase II; (2) experimental group children gain confidence through the additional practice of constructing an additional story during Phase II. Either or both of these elements can then be combined with implicit assumptions about behavior reinforcement to explain the increased hand raising observed for experimental group children, *without* any reference to expectation states or any change in the child's cognitions about his ability.

To provide empirical information relevant to the behavioral reinforcement interpretation of all these experiments we decided to conduct a set of experiments for which special controls were devised. In this set of experiments, two of the four children in each group were treated separately during Phase II. One child made up a story and received positive evaluations, exactly as the experimental group child had in all previous experiments. The other child, whom we shall call the "special control group" child, made up a story alone

with an experimenter, but did not receive the positive unit evaluations of words. The experimenter for this group was neutral, nonevaluative; the child's words were received and written down silently, and the experimenter attempted to avoid any verbal or facial expressions of either approval or disapproval. Experimenters rotated among phases of the experiment so the experimenter effects would be balanced. Experimenters were unaware of the identity of either the experimental or special control child in Phase III.

Treatment of the special control group children thus incorporates the "special attention" and "practice" features of the experimental group treatment, but omits the positive unit evaluations that the theory asserts are essential in this situation for producing expectation state changes.

Subjects for this set of experiments were third graders at the white, suburban school used for the Second Experimental Series. Table 7-10 presents the mean gain scores from the control, special control, and experimental group children.

TABLE 7-10. Average Gains in Volunteering; Special Control Group Experiments

Group	*n*	Gain
Experimental	22	1.36
Special Control	22	0.00
Control	44	−0.18

$P_{(exp>special\ control)} < 0.05$

The control and special control groups both show essentially zero changes. The experimental group shows a gain in hand raising whereas the special control group does not ($p < 0.05$).[1] The data support the predictions of Expectation States Theory and do not support an explanation based on "special attention" and "practice" arguments. It may also be noted that the special control group children did not show a marked drop in rate of volunteering, as would be expected if they had interpreted the experimenter's behavior as hostile, negatively reinforcing, or negatively evaluative. As mentioned, experimenters for this set were carefully trained to be *non*evaluative in Phase II. We conclude, both on the basis of theoretical analysis and empirical evidence, that the Expectation States Theory interpretation of the results of our experiments is most satisfactory.

[1]Recent work by Professor Barbara Sobieszek at the University of Rochester has produced results consistent with the results of this experiment. Subjects receiving differential amounts of action opportunities randomly allocated by the experimenter did *not* use this to form differential expectations for ability, even though this was the only differentiating information available to them under laboratory conditions. This result, obtained under conditions more highly controlled than ours, increases confidence in the results of the special control group experiments. We thank Professor Sobieszek for making these results available to us prior to publication.

VI. EDUCATIONAL IMPLICATIONS OF RESEARCH TO DATE AND FUTURE WORK

To place in perspective the work reported in this chapter, it must be seen for what it is intended to be: a *first step* towards applying some of the basic concepts and assumptions of the propositions of Expectation States Theory to naturalistic settings. Such attempts involve a large number of operational and procedural difficulties. To our knowledge, the only other sustained attempt to apply concepts and assumptions from Expectation States Theory for practical ends in educational research is represented by the work of E. G. Cohen and her colleagues (1968, 1970). These investigators also report difficulties with task development and experimental design like those encountered in our research program.

Despite some difficulties, it does seem possible, using relatively simple techniques, to intervene in a natural situation and to change the expectations of children so as to affect their subsequent behaviors in ways that should further educational goals. (See also Entwisle & Webster 1972, 1973*a*, 1973*b*.) Work of the Cohen group similarly suggests that it may be possible to overcome the negative effects of being black in producing differential performance expectations in mixed racial groups of grade school children.

A. RELATION TO TEACHER EXPECTANCY RESEARCH

Results of teacher expectancy studies have been disputed in some cases (see Barber and Silver, 1968) and unequivocally negative in others (Claiborn, 1969; Jacobs, 1969). The range of results of our own experiments points to possible sources of difficulty in the teacher expectancy work. First, while most children show slight increases in performance outputs even without much encouragement (the control groups), some children fail to manifest increases in performance even with extensive positive evaluations. Earlier we pointed out that white experimenters may not be able to provide very effective evaluations for black children. More recent work (Entwisle & Webster, 1973*b*) suggests that racial mismatch between adults and children is complex in its effect, interacting with social class as well.

Second, Expectation States Theory assumes a task orientation on the part of students and our experiment assumes a task where ability is equal or irrelevant. Both assumptions may be violated often in the teacher expectancy work, for many school children have little ego involvement in academic pursuits, and many already have firm ideas about their own level of ability. To the extent that children have access to objective standards or to alternative others for evaluating their performances, we would expect that the teachers' expectation for them would decrease in importance.

A third point, made by Claiborn (1969), is that there may be no chan;:s in teacher-pupil interaction—no classroom analogue of our Phase II treatment. If teachers perceive pupils to be of high potential, teachers may alter their behavior, but teachers vary in how they change their behavior to suit children's ability (Kranz, Weber, and Fishell, 1970). For example, Kranz et al., show that some teachers behave similarly towards high- and average-ability children but differently towards low-ability children. If, as in most teacher expectancy experiments, a teacher is given false reports about students' potential, she might or might not change her behaviors toward the designated children depending upon how her own behavior pattern is expressed. Also, of course, since expectations typically flow from evaluations, which in most classrooms are expressed by peers as well as by teachers, changing expectations of only the teacher may not be sufficient to produce changes in children's self-expectations. Recent work by Cohen and Katz (1972) shows that expectations of both white children and black children of a work group must be molded if black children are to improve performance.

Perhaps what is needed most at this point is more fine-grained analysis of exactly what behaviors accompany changed expectations. Our own work is one approach to this. Other approaches besides those already mentioned include Meichenbaum et al.'s observations that expectancy instructions (identification of "late bloomers") cause some teachers to increase positive interactions with students, or to decrease negative interactions.[2] Positive interactions included conveying encouragement, praise, or any attitude of satisfaction. Also Brophy and Good (1970) observe that teachers tend to praise more those children for whom they hold high expectations, and to demand more in the way of performance from them.

One of the major results of this work is the experimental task itself. It offers a means of manipulating (raising) young children's expectations without the drawbacks that affected ways of manipulating expectations used previously. The behavior influenced—frequency of hand raising—is generally considered to be important educationally. Hand raising leads to greater participation. The importance of active participation to learning is too well known to require documentation. In future reports, we hope to study further how this procedure affects racial, age, and SES subgroups of the population. Further work will extend the basic experimental design to examine the relation between expectation change and structural variables like sociometric standing.

Also, further work will study students' expectations for academic performance as a function of their own feedback over time. What happens, for

[2]In other instances when expectancy or teacher expectancy experiments have failed to achieve results that were anticipated (Goldsmith, 1970; Fleming and Anttonen, 1970) an analysis of findings in terms of Expectation States Theory may be helpful. For example, it may be that the teacher does not alter expectancy because the students do not accept the validity of the teacher's expectations.

example, to a child's high expectations for himself when he gets a low evaluation (bad report card) from a teacher? If a child has low expectations for himself, and his parents have high expectations for him, how does his classroom behavior evolve?

What is the potential of this kind of research for educational purposes? There is value first of all, in making explicit parallels between social psychological research and research in classrooms. As has been so frequently noted, there is a surprising gap between these two fields of research that hinders both. By linking classroom research to a body of pre-existing theory, one gains coherence and explanatory leverage on a whole body of research findings.

The experimental maneuver reported here may not be potent, by way of long-term educational effects, however. The present procedure, as earlier pointed out, does have the distinct advantage of avoiding the use of false test scores or of any form of outright deception. It also has the advantage of working directly on a child's expectations rather than on the expectations of some other person for the child. It is difficult, however, to imagine using a maneuver like the one described here over and over to obtain broad effects in raising children's expectations. This procedure may nevertheless have usefulness over the short term in two ways:

1. The procedure may act as a pump primer. If a child's expectations for himself are suddenly raised, as in the experiment, he may alter his actions for a short time in ways consistent with his increased expectations for himself. Any intervention within the circular series of events—improved self-expectations leading to better performance leading to improved teacher evaluations—may be effective. The short-term change in a child's actions induced by an expectation raising procedure might thus fire a chain reaction that would tend to continue once started.
2. This procedure and others that could be invented may give precise suggestions as to how teachers can convey positive expectations. Some teachers may wish to convey positive expectations but not be very adept at it. The suggestion here is that praise for one activity (word giving) leads to an increased level of another activity (hand raising) at a later date. In classrooms where many activities occur, then, the teacher need not wait, perhaps in vain, for a praiseworthy performance in arithmetic to improve a child's expectations for himself in the area of arithmetic. If the child can be encouraged for one sort of performance this may generalize to yield increased participation across the board. In fact, some situations may be used primarily to improve expectations rather than for learning per se.

The results of the Special Control Group experiments indicate quite clearly that it is the unit evaluations that determine children's expectations, or at least that by comparison with other features of the situations, such as being given action opportunities and special attention, the unit evaluations are far more important. An important extension from this is that one means that

might be thought to raise children's expectations—calling upon them more often—probably is ineffective. This conclusion may be counterintuitive in some cases, for it might be thought that if the teacher calls upon certain children more frequently this would indicate to them that she thinks highly of their answers.

However, as was shown in the Special Control Group experiment, this "special attention" or differential allocation of action opportunities is probably ineffective in changing expectations. Thus, in order to apply the results of our experiments to raising children's expectations, it is important to bear in mind that calling on the selected children more frequently will not by itself improve their expectations greatly. What is probably required is to *praise* their responses, to give them extensive positive evaluations of their performances.

To summarize results at this stage of our research, we review the three general goals adopted at the outset. First, we hoped to apply a formal theory to analyze interaction in classrooms, particularly what has been called the "teacher expectancy effect," and to show that some of the previous results in this area could be incorporated into the scope of Expectation States Theory. Second, we hoped to develop techniques of intervention that would induce specific changes in classroom interaction to improve children's learning. Third, we needed to develop a task—the story telling game—that could be used easily and effectively. The overall results of our experiments indicate some success in meeting these goals.

More specifically, the experimental results showed increases in expectations for all groups except white, middle class female third and fourth graders, and, initially, black third and fourth graders. In the case of the black children, an additional set of experiments suggested that the earlier failure to produce results may have stemmed from racial mismatch between experimenters and children (white experimenter, black children). Preliminary experiments suggest that another kind of racial mismatch (black experimenters, white children) will *not* vitiate the procedure—black experimenters may turn out to be more generally effective across all kinds of subgroups of children than white experimenters. Research with black experimenters and white children and with racially mixed (black-white) groups is now commencing. We are also observing teachers in classrooms to get information about how teachers "naturally" convey expectations, and about how other students convey expectations. The effect of the treatment was greatest for rural children, perhaps because they attend schools where few breaks in the routine occur. The schools are in remote areas, and so are seldom included in research studies or special programs. Other schools where this research was carried out are almost continually involved in activities initiated by persons not on the regular staff. Thus the urban and suburban children may be more "sophisticated" in terms of serving as research subjects and less susceptible to any attempted treatment.

Further work in this program may be organized into three general cate-

gories. First, we are working to develop additional tasks that may be used in the research. It seems desirable to develop tasks unrelated to academic activity, for example, athletic tasks and social or leisure-time tasks such as hobbies. Second, we hope to be able to study factors that govern how successful the intervention is. For example, sex of the experimenters may have a differential effect, and probably operates in rather complex ways involving cross-sex and same-sex combinations. Factors of a different kind such as the centrality of sociometric ranking of the children and their average level of performance in school may also be related to expectation raising. Third, we hope to design studies based upon other aspects of the theory. One such study springs from Assumption 3, which asserts that action opportunities and evaluations will be *distributed* in accordance with expectations held for *others*. In some cases, perhaps when children are interacting together in the absence of an adult, the assignment of an action opportunity may be equivalent to the voicing of a positive evaluation.

	Phase I	Phase II	Phase III
Control *S*'s	one story is produced (12 words); no evaluations; level of volunteering observed	control *S*'s have story read to them by another adult, from 12 to 15 control *S*'s join in one group	repeat Phase I, with same control *S*'s and experimental *S*'s as in Phase I; experimenters are rotated so the experimenter is unaware of identity of experimental *S*'s
Experimental *S*'s	one story is produced (12 words); no evaluations; level of volunteering observed	experimental *S*'s make up story individually with the same experimenter they have seen in Phase I; receive all positive evaluations; experimental *S*'s join control *S*'s at end of this phase	repeat Phase I, with same control *S*'s and experimental *S*'s as in Phase I; experimenters are rotated so the experimenter is unaware of identity of experimental *S*'s

FIGURE 7-2
Summary of experimental procedure

APPENDIX A

Formal Statement of Propositions

DEFINITION 1

A situation is task-situation S if and only if it contains:

a. at least two actors, p and o, making performance outputs;
b. an actor, e, making unit evaluations of those performance outputs;
c. no previous expectations held by p and o of their own or each other's abilities at the task;
d. task orientation of all actors;
e. collective orientation of all actors.

DEFINITION 2

e is a *source* for p in task-situation S if and only if p believes that e is more capable than p of evaluating performances.

ASSUMPTION 1

In task-situation S, if e is a source for p, then p will agree with e's unit evaluations of any actor's performances.

ASSUMPTION 2

In task-situation S, if p evaluates a series of performances of any actor, then he will come to hold an expectation state for that actor which is consistent with those evaluations.

ASSUMPTION 3

In task-situation S, if p holds higher expectations for any actor o_1 than for another actor o_2:

a. p will be more likely to give o_1 action opportunities than o_2;
b. p will be more likely to evaluate positively o_1's future performance outputs than o_2's;
c. in case of disagreement between o_1 and o_2, p will be more likely to agree with o_1;
d. p will be more likely to accept o_1 than o_2 as a source.

ASSUMPTION 4

In task-situation *S*, the higher the expectations an actor, *p*, holds for self relative to the expectations he holds for *o*:

a. the more likely is he to accept a given action opportunity and make a performance output;
b. in case of disagreement with *o*, the more likely is he to reject influence.

REFERENCES

BARBER, J. and M. SILVER. 1968. Fact, fiction and the experimenter bias effect. *Psychological bulletin* 70: part 2.

BROPHY, J. E., and T. L. GOOD. 1969. Teachers' communication of differential expectations for children's classroom performance: some behavioral data. *Report no. 25*. The Research and Development Center for Teacher Education, The University of Texas at Austin.

BROPHY, J. E., and T. L. GOOD. 1970. Teachers' communication of differential expectations for children's classroom performance. *Journal of educational psychology* 61: 367–74.

CLAIBORN, W. L. 1969. Expectancy effects in the classroom: A failure to replicate. *Journal of educational psychology* 60: 377–83.

COHEN, E. 1968. Interracial interaction disability. Unpublished research report, School of Education, Stanford University.

COHEN, E. G., M. LOHMAN, K. HALL, D. LUCERO, and S. ROPER. 1970. Expectation training I: Altering the effects of a racial status characteristic. *Technical report no. 2*. School of Education, Stanford University.

COHEN, E. and M. KATZ. 1972. Personal communication.

CRANDALL, V. C. 1963. *National society for study of education yearbook*.

CRANDALL, V. C., S. GOOD, and V. J. CRANDALL. 1964. Reinforcing effects of adult reactions and nonreactions on children's achievement expectations: a replication study. *Child development* 35: 485–97.

ENTWISLE, D. R., and M. WEBSTER. 1972. Raising children's performance expectations. *Social science research* 1: 147–58.

ENTWISLE, D. R., and M. WEBSTER. 1973*a*. Status factors in expectation raising. *Sociology of education* 46: 115–126.

ENTWISLE, D. R., and M. WEBSTER. 1973*b*. Expectations in mixed racial groups. *Sociology of education*, in press.

ENTWISLE, D. R., D. GRAFSTEIN, J. KERVIN, and M. RIVKIN. 1970. Giant steps: a game to enhance semantic development of verbs. *Report no. 81*. Center for Study of Social Organization of Schools, Baltimore, Md.: The Johns Hopkins University.

FLEMING, E. S. and R. G. ANTTONEN. 1970. Teacher expectancy or my fair lady. Unpublished paper presented at the annual meetings of the American Educational Research Association.

GOLDSMITH, J. S. 1970. The effect of a high expectancy prediction on reading achievement and IQ of students in grade 10. Unpublished M. Ed. thesis, Rutgers University, New Brunswick, N. J.

HILL, K. T. and J. B. DUSEK. 1969. Children's achievement expectations as a function of social reinforcement, sex of S and test anxiety. *Child development* 40: 547-57.

JACOBS, J. F. 1969. Teacher expectancies: their effect upon peer acceptance. Mimeograph. University of Florida, Gainesville, Florida.

KATZ, I. 1968. Factors influencing Negro performance in the desegregated school. In *Social class, race, and psychological development*, eds. M. Deutsch, I. Katz, and A. Jensen, pp. 254-89. New York: Holt, Rinehart and Winston.

———. 1970. Experimental studies of Negro-White relationships. In *Advances in experimental social psychology*, vol. 5, ed. L. Berkowitz. New York: Academic Press.

KRANZ, P. L., W. A. WEBER, and K. N. FISHELL. 1970. The relationship between teacher perception of pupils and teacher behavior toward those pupils. Paper presented at American Educational Research Association meetings, Minneapolis, Minn.

MAEHR, M. L., J. MENSING, and S. E. HAFZGER. 1962. Concept of self and the reactions of others. *Sociometry* 25: 353-57.

MEICHENBAUM, D. H., K. S. BOWERS, and R. R. ROSS. 1969. A behavioral analysis of teachers expectancy effect. *Journal of personality and social psychology* 13: 306-16.

ROSENTHAL, R. and L. JACOBSON. 1968. *Pygmalion in the classroom*. New York: Holt, Rinehart and Winston.

ROSENTHAL, R., and D. RUBIN. 1971. Pygmalion reaffirmed. Mimeograph. Department of Social Relations, Harvard University (July). Prepublication version of an invited contribution to the volume *Pygmalion reconsidered*. National Society for the Study of Education.

VIDEBECK, R. 1960. Self-conception and the reaction of others. *Sociometry* 23: 351-59.

WEBSTER, M. 1970. Status characteristics and sources of expectations. *Report no. 82*. Center for Study of Social Organization of Schools, Baltimore, Maryland: The Johns Hopkins University.

EXPECTATIONS IN MIXED RACIAL GROUPS

Doris R. Entwisle

Murray Webster, Jr.

The Johns Hopkins University

Sociology of Education, in press.

Abstract

This research studies how the status characteristics of adults and children affect adults' ability to raise a child's expectations of his own performance at school-like tasks. White adults are effective at raising expectations of white children or black children in mixed racial work groups; black adults are effective with black children but apparently not with white children in mixed groups. These results, both consistent and inconsistent with previous findings for homogeneous groups, are interpreted in light of the children's relative position in SES with respect to members of their own race. Unlike most research related to the effects of desegregation, this research examines both black children's and white children's reactions to black adults.

EXPECTATIONS IN MIXED RACIAL GROUPS[1]

One way to conceptualize educational research is by the unit of analysis. Traditionally educational psychologists have analyzed individuals, for example, how a person's perceptual skills affect his reading achievement. Sociologists of education, on the other hand, have analyzed social aggregates, addressing such issues as the school performance of subcultural or minority groups. Both lines of inquiry look at outputs, achievements of the individual or of the social group. Both tend to skip over how inputs are converted into outputs of achievement, in particular what interpersonal events, as the child interacts with his family group or with his peer group, foster achievement. If middle class children read better than lower class children, exactly what happens day by day in a middle class setting that brings about superior reading? What social processes or activities occur to account for the differences?

This paper tries to trace out sequences or patterns of social interaction that lead children to make outputs. The unit of analysis is the school child nested in a group with peers and one adult. This unit bridges both the psychological and sociological units of analysis mentioned above, and focuses on the articulation of the child with his group of significant others. The aim is to trace out how social factors and processes are translated into expectations of children and how these expectations, in turn, affect children's performance.

In the experiments reported here, adults attempt to alter children's performance expectations for themselves. In all cases adults attempt to manipulate upwards a child's performance expectations. One aim of the research is to point out what an adult can do to raise a child's expectations. Another aim is to see how status characteristics of adults and of children affect expectation-raising. Is a black adult, for example, more or less effective than a white adult in raising a black child's expectations? Does a child volunteer more often in a mixed race group than in a group where everyone's race matches his own?

The program to raise children's performance expectations developed from several sources, among them a theory describing the determinants and consequences of performance expectation states in small task-oriented groups (see Berger, et al., 1972). According to this theory, expectation states--which are roughly equivalent to beliefs about ability--are formed for actors as the result of evaluations made of their performances. Once formed, expectation states tend to persist, and they affect both the ways in which actors treat each other, and the types of actions they initiate. Some of the interaction consequences of expectation states, particularly self-confidence and willingness to engage in interaction, seem very closely related to problems encountered in educational situations.

The laboratory, however, is remote from the classroom. Producing practical results in a naturalistic setting leads to problems somewhat

different from those faced in a more highly controlled social psychological laboratory. What characteristics, for example, must be possessed by an adult in order for this adult to function as an effective raiser of children's expectations? In general we expect that an effective expectation-raiser should possess the characteristics of a "significant other" in Sullivan's (1947) language: he should be warm, trustworthy, and have the potential for affective ties to the children; he should also be perceived as highly competent to evaluate (Sobieszek and Webster, 1973). There may also be particular historical considerations which operate in educational settings, for example, the effects of racial mix of children in the school.

Self-esteem or the self-image in relation to educational performance has been repeatedly investigated (Brookover, Thomas, & Peterson, 1964; Wylie, 1963) but aside from consistent small positive correlations between self-image and school achievement, little so far has emerged to suggest how self-esteem is acquired or how it leads to varying behavior. Perhaps only when the individual is enmeshed in his network of significant others can one study what raises or lowers self-esteem. This research focuses on social events or processes that lead to small changes in children's behavior or self-view, and looks at what specific behavior by whom brings about those changes. We attempt to see how adults interacting with children can raise the child's expectations, and how social factors affect the process.

This research, besides linking laboratory research to practical concerns, also relates closely to two other current issues in educational research.

There is, first, the issue of how segregation or integration actually affects children's performance in school. Intra-classroom data on effects of integration for grade-school children are scarece. Little is known about the social-psychological factors which establish racial identity, or about exactly how social processes in mixed racial groups affect achievement. It is often assumed that blacks, because they are discriminated against and are of low status in the larger society, come to have low expectations for themselves. This set of low expectations is thought to lead to a generalized low opinion of the self (low self-esteem) as the low expectations are repeatedly confirmed. Recent findings (Beers, 1973; Rosenberg & Simmons, 1971; Soares & Soares, 1969) of relatively high self-esteem for black children, in some cases exceeding levels for whites, run counter to these suppositions and point up the complexity of integration effects.

Expectations, and expectation raising or lowering, are shaped in a social context. While adult blacks may be discriminated against, children who live in all-black neighborhoods and who attend segregated schools may be largely unaware of "outside" events. The paradoxical findings of high self-expectations for blacks may be partly explained by blacks' relative insulation, as suggested by Rosenberg and others. Data here provide evidence in support of an insulation effect for, as will be seen, black adults are uniformly effective at raising expectations of black

children while white adults are not. A child dependent upon other people like himself for evaluation may be in a cocoon where the larger society is almost irrelevant.

Of equal importance, performance expectations have both "self" and "other" components: a child's expectations for himself are defined only with regard to a specific other individual. From our perspective, it makes no sense to speak of "general low self-expectations" supposedly held by blacks. If black children have low expectations for their own performance, it is only with respect to specific other children, such as white children. A black child may have average or high expectations for his own performance when he is in a group of other black children, and then display all the behavior patterns of a child with low self-expectations when he is in a group containing white children. Our research is grounded in a theory which explicitly emphasizes the importance of social context--separation from society as a whole, and racial composition of the immediate group--in forming the child's view of himself.

Second, the issue of teacher expectancies and the self-fulfilling prophecy in educational settings has led to a shower of studies and to some controversy. Contradictory and equivocal findings have been frequent. The experiments reported here trace how teachers' expectations, the presumed "active agent" in the Pygmalion studies, may be conveyed to students. Expectations teachers hold for their students may be potent determinants of students' actual performances because teachers translate their expectations into responses that affect the child's own expectations for himself. Although there are many studies in the teacher expectancy area, none but ours as far as we know, tries to link the child's expectations into the process. The present experiments focus directly upon the child's expectations, which may be the most important part of the "Pygmalion effect." If a teacher expects more from a child, he gives evidence of his high expectations for the child. The child, besides producing more in response to the teacher's demand, probably changes his view of himself so that it conforms with the teacher's view of him. In other words, something cognitive and something persistent occurs in the child when the teacher gives evidence of high expectations.

A child builds performance expectations for himself on the basis of responses supplied him by significant others. These expectations then persist as part of his ability self-concept. We are investigating how children actually form expectations about their own ability to do school work, and when these expectations shift as a consequence of actions taken by others. What set of persons--parents, peers, teachers, others--constitute the set of significant others for this process? Putting the child in the loop, as the present research tries to do, but which other teacher expectancy studies fail to do, allows a number of related issues--the effect of praise, reinforcement properties of adults-- to be integrated into a consistent conceptual framework. Further discussion of these relationships will follow presentation of the experiments.

Previous Work

In earlier studies (Entwisle & Webster, 1972, 1973, 1974) procedures were devised for raising children's expectations for their own performance, and for improving their willingness to participate in a naturalistic classroom task. An adult works with one child selected from a four-child group and follows a prescribed pattern to raise that child's expectations (see below). Experiments so far, with one exception, have been homogeneous with respect to race; that is, groups of black children or white children met with adult experimenters who were of the same race as the children. White children from grades one through four and from both rural and suburban backgrounds generally responded significantly to the expectation-raising treatment administered by a white adult. Black adults with black inner-city children were also effective. White adults with black inner-city children were the single exception in terms of racial matching and this combination was relatively ineffective.

Research reported in this paper extends the previous work by studying several kinds of racial mixing. Some experiments focus on black adults working with white suburban and white rural children's groups. Other experiments deal with mixed racial children's groups, where both white and black adults work with groups where two children are white and two children were black. After reviewing the experimental procedure, we will present separately results for these two new lines of research, (i) black adults with white children's groups, and (ii) black or white adults with integrated children's groups.

The Procedure

Children in second, third, and fourth grades who attended schools in the greater Baltimore area served as subjects. Middle-class young adult women, some black and some white, served as experimenters. At the outset the children met together and groups containing four children, all of the same sex and grade, were formed and designated as "Team 1," "Team 2," etc. Children were told that their team was about "to play a game making up stories," and that the teams would be competing for a prize: everyone on the winning team would get a prize. Insofar as possible, members of a team came from different classrooms. The teams then went to separate rooms, with one experimenter accompanying each team.

The experiment consists of three phases. (See Entwisle & Webster, 1972 for a more detailed description of the experimental procedure.) In Phase I each team supplies 12 words to fill in a story skeleton. The story skeleton consists of incomplete sentences: "Once upon a time there was a ____." Children are instructed to raise their hands if they can think of a word(s) to complete the sentence. The experimenter cautions the children not to raise hands to volunteer a word unless "you are sure you have a very good word--one that will help the team score." Given this emphasis on task performance, hand raising is taken as an operational measure of each child's level of performance expectations for himself. The experimenter notes unobtrusively how often each child volunteers

(raises his hand), then chooses one child to give the "team's word," being careful to choose each child equally often as a respondent. The experimenter is businesslike and does not praise or otherwise reinforce or evaluate responses in Phase I, merely noting them down.

In Phase II one child (whose rate of volunteering in Phase I is near the median for the group), is selected for the expectation-raising treatment. In Phase II the selected child makes up a story by himself using a new story skeleton. Every word he gives, as well as the overall story, receives strong praise and positive evaluation from the experimenter--a treatment which our analysis indicates (see Entwisle & Webster, 1972, 1973) should raise his performance expectations, and later, his rate of group participation. Note that the treatment to raise the child's expectations involves praise of responses, not of hand-raising; when a child is completing a story by himself he does not, of course, need to volunteer.

In Phase II the three children not selected from each team for the expectation-raising treatment leave the room where Phase I is carried out. These untreated children of each team constitute a control group, and spend Phase II in the central room listening to a story being read. (Other experiments (Entwisle & Webster, 1974) have established that this story-listening control treatment is equivalent in its effect to a control treatment in which an individual control child fills a story skeleton with an experimenter but the experimenter remains neutral. This equivalence points to the raising of expectations, as a consequence of positive feedback from the experimenter, as the crucial part of the Phase II treatment rather than to such things as telling a good story, isolation with an adult, or other incidental accompaniments of the Phase II treatment.)

Phase III is a repetition of Phase I. The original teams reassemble in the experimental rooms and use a third story skeleton to produce a new story. The experimenter again remains neutral, calling upon each child an equal number of times, and notes how many times each child raises his hand. Experimenters are rotated between Phases II and III so that during Phase III they do not know the identity of the child who received the expectation-raising treatment in Phase II.

Results

The analysis focuses upon changes in the rate of hand raising. The question at issue is whether the child who received the Phase II treatment raises his hand more in Phase III than in Phase I. His change between Phase I and III is compared to changes between Phase I and III for a selected child who received the control treatment in Phase II.

A word is needed about the choice of an individual control child whose gain is used as the baseline against which to measure gain by the

experimental child from the same group. For this work a stringent selection procedure, which entails considerable exclusion of data, was adopted for several reasons. First, group climate is probably important in its impact on volunteering, so children from the same group are matched in the analysis. Second, close Phase I matching of experimental and control children was undertaken to guard against biasing of results in favor of predictions. Biasing can occur in three ways: (a) through the mean score for three control children tending to exceed the average score of experimental group children in Phase I, because although children occasionally volunteer 11 or 12 responses in Phase I, it rarely happens that a child volunteers once or not at all; (b) through ceiling effects when scores in Phase I are close to the maximum possible score (12); or (c) through regression effects, where disparate Phase I scores might lead to regression effects either through the experimental child's score regressing upward if his score in Phase I is relatively low, or through the control child's score regressing downward if his score in Phase I is relatively high.

For these reasons, the following rigorous procedure for screening experiments was adopted. First, a particular control child's score, rather than the mean score of three control children in Phase I, was taken as the baseline. The two children are participants in the same experiment. Second, only experiments where experimental and control children are closely matched in Phase I are included for analysis. The criteria for close matching are (a) unless the experimenal child's Phase I score and the control child's Phase I score can be matched within two units or less, the experiment is discarded (regression effects[2]); (b) if the Phase I score for the experimental child or the child selected to be his control is 11 or 12, the experiment is discarded (ceiling effects).

Experiments with Black Adults and White Children's Groups (Suburban and Rural)

One set of experiments involved black adults with white rural third-graders. The black adults were middle-class black women students at a private university. The children lived in a farming area about 30 miles north of Baltimore City and attended an all-white school. The mean gain in volunteering between Phase I and Phase III for treated children over gain for control children in 17 experiments is 1.94, significant well beyond the 5% level.

Another set of experiments involved the same black adults with white suburban third-graders. These children attended an all-white school located just over the city line from Baltimore in a lower middle class area. The mean gain in volunteering for treated children over gain for control children in 28 experiments is 1.66, also significant beyond the 5% level.

When data for black adults and white suburban or white rural children are combined with earlier data (Entwisle & Webster, 1972, 1973, 1974) a 3 X 2 factorial design can be manufactured. This design is comprised

of third graders drawn from three residential loci treated by experimenters of two races, black or white. Table 1 summarizes the mean gain of treated children compared to control children for the various combinations of experimenters and children.

The reader should be warned that there are procedural differences among the experiments. The experiments summarized in Table 1 were carried out over a 5-year period. Both male and female experimenters were used in some of the early work, and experiments were aggregated in the earlier analyses. Nevertheless except for experiments with black inner-city third-graders and white experimenters,[3] all combinations produce fairly sizeable gains in expectations of treated children compared to children given a control treatment. In all these experimants groups of children were homogeneous with respect to race, so major interest attaches to the effects of matching or mixing of race between adults and children. As mentioned, all combinations are effective except the combination of white adults with black inner-city children.

Experiments with Black Adults, White Adults, and Racially Mixed Groups of Children

In another set of experiments, racially mixed groups of children participated in a procedure identical to that described above. Each group contained two white and two black children. These racially mixed children's groups met with white experimenters or black experimenters. As will be explained, a white child or a black child was selected to receive the expectation-raising treatment, and the remaining three children, two of one race and one of the other, received the control treatment.

The child chosen for the expectation-raising treatment in Phase II either matched or differed from the experimenter in terms of race. This gives four possible types of experiments (a 2 X 2 factorial design based on race-of-experimenter and race-of-child) with black experimenter and a black treated child, black experimenter and a white treated child, white experimenter and a black treated child, and white experimenter and a white treated child.

An integrated school with about 60% black and 40% white children in a lower-class Baltimore neighborhood provided subjects. Carrying out the experiments led to exhaustive sampling of white children enrolled in grades 2, 3, and 4 and to an 80% sample of black children in those grades.

In advance it was difficult to predict what would happen with racially-mixed groups of children. Perhaps in mixed groups expectations of black children would be difficult to raise. As far as race of adults is concerned, black adults had proved effective with white rural and white suburban children (see above) but white adults had been ineffective as expectation raisers for black inner-city children. The same criteria for screening experiments mentioned above led to selection of 29 experiments with mixed racial groups. Ceiling and regression effects should be minimal for these 29 experiments.

The gains in volunteering for treated children compared to gains for control children in 29 experiments are presented in Table 2. For example, in 14 experiments where the experimenter was white and the treated child was black, the treated child on the average gained two more units between Phases I and III than the control child. Other mean gains listed in the first row of Table 1 show analogous comparisons for other combinations of experimenter's race and child's race. One-sided t-tests appropriate for matched groups show significant differences favoring treated children for three experimenter-subject combinations (white E with white S's, white E with black S's, and black E with black S's), but not for black E and white S's. Variances associated with these means are relatively homogeneous.

Further examination of the data for the black E-white S groups, where there is not a significant gain, was undertaken. If an alternate set of criteria for selecting experiments is used, a few more experiments can be included in this condition to further check the conclusion. Accordingly, if experiments are selected where matching of control and experimental children is not within 2 units but where ceiling effects are guarded against and where instead the control child and experimental child are matched for race (11 groups), the mean gain score drops to near zero (.09). These results increase our confidence in the "no difference" conclusion for this condition. These same alternative criteria for selecting experiments when used for the white E-black S condition lead to confirmation at a higher level of confidence of the previously established finding of a gain for experimental children.

The matched-race experiments (white E-white S and black E-black S) show statistically significant gains, even though the number of experiments is small. The findings for matched-race combinations observed here are consistent with previous findings for matched-race E's and S's where members of teams were homogeneous with respect to race (see Table 1). Further experiments could not be run because the subject pool was used up.

Table 3 shows the mean initial (Phase I) rank in group of experimental and control children. In all groups, white and black children volunteer about the same amount in Phase I, with black children volunteering only slightly more than white. This table suggests that black and white children's performances are initially much the same.

Discussion

Status Difference and Status Distance

Results from the present experiments, together with earlier experiments reviewed above, make a complicated picture. To review the findings: (1) when the race of the experimenter matched the race of the treated child and children's groups consisted of a single race, significant increases in the expectations of the treated child occured. Black experimenters have been uniformly effective in raising the expectations of black children, and white experimenters have been uniformly effective with white children. (2) In the one case previously reported (Entwisle &

Webster, 1974) where race of experimenter differed from race of children, white adults were ineffective with all-black inner-city children's groups. The findings of the present experiments show (3) significant increases for black adults with all-white groups of rural and suburban children, (4) but no significant effect for black adults interacting with white children drawn from mixed racial groups even though (5) white adults are effective with black children drawn from racially mixed groups. Effects of racial mismatch are apparently complex. Two possible explanations are suggested, one based on racial mismatch and the other on social class.

Mixing of races between experimenters and children may make children less likely to see the adult as a credible source. The black child encouraged by the white experimenter may feel he is being implicitly compared with other white children even though the children surrounding him are all black (finding 2 above) or a white child encouraged by a black experimenter may feel he faces "stiff competition" when he is in an integrated group led by a black experimenter (finding 4 above). But this explanation is not entirely satisfactory; it is hard to see why this same explanation would not apply equally well to the white child in an all-white group encouraged by a white experimenter (finding 3 above) or to the black child encouraged by a white experimenter, who competes against white children in an integrated group (finding 5 above).

Supplementary evidence also weakens the racial-mixing explanation. In the experiments where children's groups were racially mixed an analysis of variance performed on rates of volunteering in Phase I with race-of-experimenter and race-of-children taken as two fixed-effect factors, shows no significant main effects or interaction effects. Children of both races volunteered about the same amount whether the experimenter was white or black. In Phase I white children volunteered somewhat more for black E's (7.71) than for white E's (6.71), although not significantly more in 28 and 26 experiments respectively. This suggests that initially at least, white children respond at about the same level to both types of experimenters, and, if anything, are somewhat <u>more</u> responsive to black adults. This is confirmed in Table 3.

Social class or SES level is a major point of difference between groups of children in this and earlier studies. Altogether four residential loci are involved: rural, suburban segregated white areas, segregated inner-city black areas, and an integrated area located at the black-white interface in Baltimore. Blacks in the integrated area are considerably higher up the SES ladder than blacks in completely segregated ghetto neighborhoods in east and west Baltimore where the studies with all-black children's groups were carried out. On the other hand, whites living in the integrated area in Baltimore City where the racially-mixed groups were secured are much below the SES level of the rural or suburban whites, being at the low extreme of white SES in Baltimore.

This suggests that children of very low SES relative to others of their own racial group--whites from integrated neighborhoods or inner-city blacks living in the ghetto--may perceive adults of the opposite race as hostile. Opposite-race adults, if they are perceived as hostile, would not be accepted as expectation-raisers by such children.

Precisely why low SES children of both races should be most likely to perceive hostility from other-race adults is not clear, although Katz (1968, 1970) reports that white adults are seen as hostile by black children. From the present experiment it may be possible to refine Katz' idea as follows: "Children of very low status in any group are apt to perceive adults as hostile if adults belong to a different racial group from the child." Our complex results can be interpreted in a single framework if this explanation is used.

Social class or residential locus is a second parameter in addition to race which leads to social distance between experimenters and children. Perhaps to achieve credibility as an expectation-raiser the source must be far enough removed to be perceived as competent to judge, but not so far removed that he recedes as a significant other.

These experiments, carried out over a long time span, include some differences among experimenters that may affect the findings. The black experimenters who were effective with all-black inner-city children's groups were lower middle class, considerably lower in SES than the white experimenters who worked with similar groups at the same time. There was, in addition to matching on race, less social class distance between experimenter and children in the case of black E's and black ghetto children than for white E's (upper middle class). The black experimenters who were ineffective with white children in mixed groups, on the other hand, happened to be upper middle class blacks, so that again race and social class were strong points of difference between adults and children.

Racial Matching

A number of studies outside the teacher expectancy field examine interactions between race of children and race of experimenter when an adult reinforces (praises) or criticizes the child in simple performance tasks (see e.g. Allen, Dubanoski, & Stevenson, 1966). In most of these studies social class is not studied and sometimes the child's task is ambiguous. Nevertheless in general a same-race effect is found: children perform better, faster, or more effectively when the adult administering reinforcement is of the same race as their own. In our experiments to date (which involve hundreds of children in both segregated and integrated settings) we have found no exceptions to the effectiveness of racial matching over a wide range of children's social class and age levels. In every experiment when adults and children are of the same race the adult is able to raise children's expectations. Interestingly, some recognition of possible advantages of racial matching in college tutoring programs occurs in practice. In a survey of 46 colleges having peer-tutoring programs, most programs with a large percentage of ethnic minority students have tutors from the same ethnic group (Reed, 1973).

School Integration

It is tempting to speculate about the relevance of the experiments reported here to problems and tactics of school integration. Naturally

such speculation should be guarded since the results here, although statistically significant, are based on relatively small, and perhaps unrepresentative samples. Particular schools in the Baltimore area may not be good barometers of social climate in other schools in the same area, let alone elsewhere. Also the experimental situation is organized around a single activity (story-telling), a single behavior (volunteering), and a single cognitive change in the children (expectation-raising), whereas teachers in classrooms work with larger groups on many topics where cognitive changes involve learning rather than only expectation-raising. Nevertheless since little is known about how school integration affects school performance, particularly for young children, some extrapolation may be fruitful.

To amplify a point suggested earlier, a conception of childrens' self-image which does not include reference to a particular social situation is misleading and over-simplified. A child holds performance expectations for himself relative to some other individual or individuals, performing a known task. If a black child holds low self-expectations, it is because he has been unable to perform well (received negative evaluations), at a particular task, or because he is in a group of others whom he has reason to suppose can perform better than he. Experiences of past success and failure <u>by comparison with others in a group</u> determine expectation states; if the relative amount of positive evaluation given a child changes, or if he moves from one group to another, his relative self-expectations and expectation-related behaviors will change accordingly. Recognizing the relative nature of expectations, as well as of the evaluation process by which they are built up, should permit intervention of several sorts to combat social sources of unrealistic low self-expectations.

It is often speculated that children of different races in integrated schools are likely to regard each other with suspicion or hostility. One effect of such hostility might be to depress voluntering rates in the experiments reported here, especially rates of black children. From another point of view, if children perceive race as a relevant status characteristic, they will form expectations for each other based upon their respective races (see Berger it al., 1972). The effect of this process also would be to depress expectations black children hold for themselves--especially in Phase I before the experimental treatment intervenes--and thus, their volunteering rates.

The fact that black children <u>do not</u> have lower volunteering ranks in Phase I than whites in any of the experimental conditions suggests that neither of these processes occurred. There is no evidence that racial tensions between children depressed blacks' volunteering in Phase I, so children in this study may not perceive race as a relevant status characteristic in forming performance expectations for themselves and each other. Other data (presently unpublished) for older children in this same school reveal that both white and black children have many fireds in school of the opposite race (even though they do not have cross-race friendships in their neighborhoods) and have no preferences in terms of race as far as "next year's teacher" is concerned. Surprisingly, one class of fifth-graders in this school was unaware that blacks constitute a numerical minority group until this information was told them by the teacher as she taught a social studies unit in intergroup relations. More research would be needed

to investigate these suggestions systematically, but the fact that this school has remained successfully integrated for several years perhaps has contributed to breaking down some "interaction disabilities" based upon race. Whether the same climate would prevail in a newly-integrated school is not known.

Related to this point, Cohen and Roper (1973) tried to increase performance expectations of black children participating in biracial work teams. Junior high aged children of both races from segregated backgrounds were brought together for experiments, and apparently all children formed low expectations for blacks and relatively high expectations for whites on the basis of the race status characteristic. To overcome effects of race in order to raise blacks' performance outputs, the expectations of white children for black children had to be changed as well as expectations of black children for themselves. Extraordinary efforts were necessary to accomplish this, documenting the strength and the persistence of racial bases for performance expectations. Black children in the present experiments came from an integrated background and apparently neither they nor their white classmates held low expectations for blacks' performance.

History of previous contact, particularly equal-status contact between schoolchildren across races, and age differences, may best explain the difference among findings. Elementary school black children like those in the present study may have higher self-concepts than older children who come from segregated schools or who have been compared unfavorably to whites. The same general effects would also help explain the complex interactions found between race of adult, race of child, and child's social class level in this series of experiments and previous experiments of the same type.

Most previous research on performance in mixed racial groups has involved college students. There is very little research on mixed racial groups involving children of elementary school age, or on desegregation effects for younger children (see Cohen, Pettigrew, & Riley, 1972). The work reported here may have some relevance to desegregation strategies in two respects.

First in a school which has been desegregated for longer than any of its present students have been in it, these data show an important finding by the absence of evidence that performance expectations are based upon race. One goal of school desegregation is to make race non-relevant as a sufficient basis for forming performance expectations. There is no evidence in our data that these children formed expectations for one another on the basis of race. By comparison with other results with children from segregated schools, these data are encouraging. Comparing these results of ours to results Cohen and Roper (1973) report may well indicate that sustained integration of an elementary school can break down the use of race as a basis for performance expectations. This constitutes what is to our knowledge one of the very few quantitative indicators of positive effects of desegregation.

Second, discussions of desegregation effects concern mainly reactions of blacks in schools where some classmates and teachers are white (see, e.g., Katz, 1968). Issues involving white children's reactions to black teachers remain virtually unexplored. In racially mixed classes, for example, will a black teacher or a white teacher be more effective? Are there beneficial effects in terms of positive intergroup relations from having first- and second-graders in white segregated neighborhoods taught by black rather than white teachers? Or what are the effects upon inner-city black children of white middle-class teachers? Compared to the efforts necessary to bus children between schools, it would be relatively easy to "bus" teachers. This aspect of school integration, the deployment of the black or white teacher, although easily manipulable, has been little investigated. Most studies so far concern "racial balance" where racial balance has to do only with the racial mix of students. The present research begins to broaden research along the teacher dimension and suggests what factors--such as status distance and SES matching--interact to affect teacher effectiveness as a significant evaluating other.

Teacher Expectancy

The research reported here is also closely related to research in the teacher-expectancy field but differs in two crucial respects: it is directly concerned with the child's expectations and it attempts to specify what actions taken by adults affect children's expectations. In other work in this field, much of which is equivocal, teachers' expectations are manipulated and then the subsequent behavior of children, usually performance on standardized tests, is observed. As we have pointed out in detail elsewhere (Entwisle & Webster, 1973b) there are many ways to explain previous findings: teachers may not respond to the expectation-raising maneuver because they do not believe it or do not remember it; teachers may not change their behavior in ways the student perceives as manifesting positive expectations; students even if made the target of the teacher's positive expectations may not respond.

A number of investigators have begun to look at the find-grained behavior of teachers in relation to students to try to pinpoint what actually occurs in student-teacher interaction in the classroom (see e.g., Brophy & Good, 1970). These studies, with one exception, do not examine race or social class. However, Rubovits and Maehr (1973) studied 66 white female teachers-in-training in interaction with junior high school students with the students in 4-person groups (2 black children and 2 white children). Random members of each racial pair were labelled as "gifted" and "non-gifted." It turned out that teachers called on white students more and praised them more than black students. There was, in addition, a startling "Race X Label" interaction. Although all blacks were given less attention and less praise than whites the gifted blacks were given the <u>least</u> attention and praise, even less than the non-gifted blacks.

The Rubovits and Maehr study examined "teachers" (actually persons like our experimenters, female college students) in free interaction with racially-mixed students' groups where the adults could allocate action opportunities and then respond to children's actions. The Rubovits and

Maehr study cannot be directly compared with our experiments for several reasons, one being that children in a "small midwestern city" are probably not comparable to our S's in terms of social class. For another, our experiments have a tightly prescribed protocol for adults' actions, especially in that the number and type of action opportunities allocated cannot vary across race of the child. On the other hand, their observations do suggest why white adults may not be credible sources for black children in some cases, for black children may have had a previous history of relatively unfair treatment by white adults like that seen in the Rubovits study.

We are presently conducting further work on children's expectations to see (1) whether expectations generalize from one task to another and (2) what kinds of actions taken by children may affect expectations of other children.

Footnotes

[1]This research program is supported by Office of Education grant OEG-3-71-0122. We gratefully acknowledge this financial support. Students and teachers at Fifth District Elementary, Loch Raven Elementary (Baltimore County) and School 53, Baltimore City, cooperated in this research. We are grateful to them and to their principals, Mr. Frank Tondrick, Dr. Evart Cornell, and Mr. Elliot Epstein. We are also grateful to Dr. George Gabriel, Director of Research, Baltimore County, for his help.

[2]The closest of possible matches was taken when more than one control child met the 2-unit separation criterion. If two control-group children are tied in terms of closeness to the experimental child, the control child whose race matches the experimental child's is selected.

[3]Black inner-city fourth-graders and white experimenters showed an even smaller gain, an average increment of only 0.11 for 16 experiments.

References

Allen, S. A., R. A. Dubanoski, and H. W. Stevenson.

1966 "Children's performance as a function of race of E, race of S, and type of verbal reinforcement." Journal of Experimental Child Psychology, 4 (November): 248-256.

Beers, J. S.

1973 "Self-esteem of black and white fifth-grade pupils as a function of demographic categorization." Unpublished paper presented at the annual meetings of the American Educational Research Association, New Orleans, La.

Berger, J., B. Cohen, and M. Zelditch.

1972 "Status characteristics and social interaction." American Sociological Review, 37 (June): 241-255.

Brookover, W. B., S. Thomas, and A. Peterson.

1964 "Self-concept of ability and school performance." Sociology of Education, 37 (Spring): 271-278.

Brophy, J. E., and T. L. Good.

1970 "Teachers' communication of differential expectation for children's classroom performance." The Journal of Educational Psychology, 61 (October): 365-374.

Cohen, D. K., T. F. Pettigrew, and R. T. Riley.

1972 "Race and outcomes of schooling." Pp. 343-370 in F. Mosteller and D. P. Moynihan (eds.), On Equality of Educational Opportunity. New York: Random House.

Cohen, E. G., and S. Roper.

1973 "Modification of interracial interaction disability: An application of status characteristic theory." American Sociological Review, 37 (December): 643-657.

Entwisle, D. R., and M. Webster.

1972 "Raising children's performance expectations." Social Science Research, 1 (June): 147-158.

1973 "Status factors in expectation raising." Sociology of Education, 46 (Winter): 115-126.

1974 "Raising children's expectations for their own performance: A classroom application." Chapter 7 in J. Berger, T. Conner, and M. H. Fisek (eds.), Expectation States Theory: A Theoretical Research Program. Cambridge: Winthrop Publishers.

Katz, I.

1968 "Factors influencing Negro performance in the desegregated school." Pp. 254-289 in M. Deutsch, I. Katz, and A.Jensen (eds.), Social Class, Race, and Psychological Development. New York: Holt, Rinehart, & Winston.

1970 "Experimental studies in Negro-White relationships." Pp. 71-118 in L. Berkowitz (ed.), Advances in Experimental Social Psychology, Volume 5. New York: Academic Press.

Reed, R. J.

1973 "When students teach: Tutoring by peers." The Research Reporter, 8 (1): 1-4. The Center for Research and Development in Higher Education, University of California, Berkeley.

Rosenberg, M., and R. G. Simmons.

1971 Black and White Self-Esteem: The Urban School Child. Washington, D.C.: The American Sociological Association Rose Monograph Series.

Rubovits, P. C., and M. L. Maehr.

1973 "Pygmalion black and white." Journal of Personality and Social Psychology, 25 (February): 210-218.

Soares, A. T., and L. M. Soares.

1969 "Self-perceptions of culturally disadvantaged children." American Educational Research Journal, 6 (January): 31-35.

Sobieszek, B., and M. Webster.

1973 "Conflicting sources of evaluation." Sociometry, 36 (December): 550-560.

Sullivan. H.

1947 Conceptions of Modern Psychiatry. Washington, D. C.: W. H. White Psychiatric Foundation.

Wylie, R. C.

1963 "Children's estimates of their schoolwork ability, as a function of sex, race, and socioeconomic level. Journal of Personality, 31 (June): 203-224.

Table 1

Mean Gain Between Phase I and Phase III, Experimental Children over Control Children (number of experiments in parentheses)

	Children's Residential Loci		
	Suburban White	Inner-City Black	Rural White
White Experimenters	1.58* (14)	0.97 (20)	2.03 (20)
Black Experimenters	1.66 (28)	1.45 (19)	1.94 (17)

* For boys only. The treatment was ineffective for girls. See Entwisle and Webster, 1973b, for details.

Table 2

Gains in Volunteering for Mixed Racial Groups
(number of experiments in parentheses)

	White Experimenter		Black Experimenter	
	White Treated Child	Black Treated Child	White Treated Child	Black Treated Child
Mean Gain	4.16 (3)	2.00 (14)	1.00 (8)	1.75 (4)
t-values	2.93*	2.53**	0.97	1.70*

* $p < .05$

** $p < .01$

Table 3

Volunteering Ranks for Mixed Racial Groups
(rank 1 high; rank 4 low)

	White Experimenter		Black Experimenter	
	White Treated Child	Black Treated Child	White Treated Child	Black Treated Child
Average Initial Rank in Group				
Exper. Child	3.00	2.69	3.25	2.62
Control Child	2.83	2.40	3.00	2.62

RAISING EXPECTATIONS INDIRECTLY

Doris R. Entwisle

Murray Webster, Jr.

The Johns Hopkins University

Abstract

Previous experiments have demonstrated that it is possible to raise children's expectations for tasks like those children perform in classrooms. Experiments reported here show that children will "generalize" expectations from one task to a second dissimilar task. If children's performance at a task involving planning a meal are positively evaluated, they subsequently demonstrate that they have raised their expectations for their own performance at a story-telling task. Results are consistent with previously reported data in that the generalization effect, while significant, is smaller in magnitude than the direct effect where story-telling is used both for positive evaluations and measuring expectations.

RAISING EXPECTATIONS INDIRECTLY[1]

The classroom presents a continuous series of unfamiliar tasks to each child. At particular times, for instance, children must learn grapheme-phoneme correspondences, cursive handwriting, the multiplication tables, how to read a map, or to ask to get a drink of water. Considerable evidence now exists that how well a child expects to do at each new task affects his performance, his self-confidence, his interaction patterns with teachers, his interaction with other students, and often his level of learning (see, for example, Brophy & Good, 1970; Entwisle & Webster, 1974b; Rosenthal & Jacobson, 1968).

For some time we have been engaged in a series of experiments designed to analyze determinants and interaction consequences of children's expectations for their own performances. One aim of this research is to produce simple techniques to allow intervention in the chain of social psychological events involved. We believe that children develop relatively enduring conceptions of their own likely success at specific tasks as the direct result of positive or negative evaluations received from the teacher and from other students (Entwisle & Webster, 1972a, 1972b, 1972c). Further, one can intervene in a classroom-like setting to produce high self-expectations by giving selected children positive evaluations of their past performances at problems, and these induced high self-expectations will be reflected in increased frequency of hand-raising in response to group-directed questions (Entwisle & Webster, 1974b). Finally, we have developed a simple procedure for expectation raising which has been successful with children of varying demographic characteristics (age, sex, and race, in urban, suburban, and rural residential locations (Entwisle & Webster, 1972d, 1973, 1974a)).

The basic experimental task consists of three phases in which groups of four children work to make up stories. The experimenter reads an incomplete sentence, and where words are missing, pauses to let children think up a "good" (appropriate and interesting) word. Each child who thinks he has a good word raises his hand, and one of them is called on to supply the "team's word" for that sentence. Raising the hand in response to group-directed questions is taken as the operational measure of each child's level of expectations for success at this task.

Phases I and III are measurement phases, in which children's expectations are recorded. During phase II, an attempt is made to raise one child's expectations by having him complete a story entirely by himself, filling in all the needed words when the experimenter gives the sentence cues. (Other children, who form a control group, meet elsewhere with another experimenter.) The crucial aspect of the experimental treatment is that each response the child gives is positively evaluated. In the measurement phases, no evaluations are given. By comparing the phase I to phase III gain scores on hand-raising frequency for experimental group children (those receiving the phase II treatment) to scores for matched control group children (who do not receive any evaluations of

performances), one can determine if the experimental treatment led to increased rates of hand-raising.

Experiments have so far involved about 1000 children, and have been widely effective across sociological subgroups. They show it is possible to raise expectations through controlling performance evaluations, and these improved expectations are reflected in a behavior widely regarded as an important determinant of actual learning, willingness to engage in problem solving attempts in the classroom. But whether one can raise expectations for a new task has not so far been tested. The research to be reported here involves such a test.

Recent theoretical work (Berger & Fisek, 1974; Kervin, 1973) asserts that in the absence of information regarding competence at a specific task, individuals will form expectations from knowledge about some other task. In other words, if an individual needs to decide how likely he is to perform well at some new task and if there is no available information regarding his performance at task A, he will use knowledge of his ability at task B and predict that his ability at task A is the same as that at task B. Laboratory experiments with college students support this claim. In the classroom, then, if a child has no idea how well he can do some new task such as arithmetic, but knows he is good at spelling, he will decide as a "best bet" that he can also do well at arithmetic.[2]

Whether the process works as simply as just described, particularly whether expectations will generalize across very different tasks and whether grade school children in a classroom will behave like college students in a laboratory, remain undemonstrated. We therefore undertook a set of experiments to shed light on the expectation generalization process with grade-school children.

The experiments reported here were designed to resemble as closely as possible previous experiments in order to permit direct comparison of results across experiments. As mentioned, the experiments consist of three phases. In phase I a group of four children are assembled and the story-telling task is described to them. Then they build a story by volunteering 12 words, exactly as in previous experiments. In phase II, one child is selected to receive the expectation-raising treatment at the unrelated task, and the other three children go to another room and have a story read to them. Instead of making up his own story as in previous research, the experimental group child works on an entirely different task, one called "meal-planning." The meal-planning task is presented to the child as "choosing foods which are nutritious and go together well for a holiday dinner."[3] A large bulletin board containing colored pictures of various kinds of food is placed before the child and he selects food pictures one at a time and places them together on another bulletin board nearby. The evaluation procedure is comparable to that of previous experiments: the experimenter, after every food choice, tells the child he has made a good and nutritious choice, and the resulting dinner after all choices have been assembled gets a very high overall evaluation. Then the original groups are reassembled for phase III, experimenters are rotated between groups to prevent their knowing which child received the expectation-raising treatment, and the story-telling task is repeated. Nothing is said about possible relevance of the two tasks, meal-planning and story-telling.

Subjects for these experiments were third and fourth graders from two schools enrolling only white children in rural Maryland, and one school in Baltimore City enrolling only black children. Children of both sexes participated in the experiments, though each four-person experimental group was composed of only one sex. Experimenters were white women for the rural children and black women for the black children.

If expectations generalize in this context, one would predict two results. First, by comparison with the untreated control group children, the experimental group children should show greater phase I to phase III gains in expectations, measured by rate of volunteering. Second, because tasks differ between phases expectation-raising effects should be weaker than effects previously observed when story-telling was the task in all three phases.

Table 1 presents the average gain scores for experimental and control groups in both settings. Experimental group children (white) in rural Maryland show a mean gain of 1.44 between phase I and phase III; control group children show a comparable gain of .33. The difference or "expectation advantage" for experimental group children is +1.11, and is significant at about the .03 level using a one-sided t-test for matched pairs.[4] Similarly experimental group children (black) in Baltimore show a mean gain of 1.10 between phase I and phase III; control group children show a negative gain of -0.45. The difference 1.55 is significant at about the .02 level.

Table 2 presents comparable data for previous expectation experiments when evaluations of story-telling were used to raise expectations. As may be seen, although both procedures produce statistically significant results, the effect of indirect expectation-raising is not as large as the effect of direct manipulation in previous experiments. In previous experiments (Entwisle & Webster, 1973) the mean gain for third grade white rural children was 2.60; for third grade black inner city children it was 1.85.

The gain scores in Table 1 show that it is possible to raise children's expectations for the story-telling task by giving them positive evaluations for an unrelated task, planning a meal. The expectation generalization process apparently occurs for young children in settings resembling those of a classroom. The results are consistent with results in more highly controlled laboratory settings with college-age subjects.

In practical terms these experiments show that an indirect expectation-raising procedure may operate in ordinary classrooms where it is not possible to control evaluations directly. A teacher need not wait, perhaps in vain, for a praiseworthy performance in arithmetic: giving a child heavy positive evaluation in some other area probably affects his general expectations, and thus his willingness to try arithmetic problems.

The two tasks used for these experiments were dissimilar, and this suggests that two tasks need not be perceived as relevant in order for generalization to occur. In laboratory experiments on expectation generalization the two tasks were also dissimilar: matching English and non-English word meanings in one case, and estimating the area of colored figures in the second. By contrast, learning experiments on stimulus

generalization use closely related stimuli, such as a 600 cps and 700 cps tone. The expectation generalization process assumes a cognitive linking of two stimuli (ability beliefs) which are not at all intrinsically similar.

In further experiments we hope to test further the applicability of the generalization phenomenon across sociological subgroups of the population, and in conditions of explicit relevance of the two tasks.

Footnotes

[1]This research was supported by Office of Education grant OEG-3-71-0122. We thank Frank Tondrick and Charlotte Harper, principals of the schools involved in the experiments, and students and staff in these schools for their help with the research.

[2]The theory says that generalization will occur in all cases except where the individual knows for sure that the tasks require unrelated skills. In other words, the burden of proof is upon a claim that abilities are unrelated, and it is not necessary for the individual to know for sure that the tasks are related.

[3]We thank Lynne Roberts of the University of Washington, Seattle, for her assistance in developing the meal-planning task.

[4]A word about computational procedures is in order. In these experiments there are possible sources of spurious successful results; for instance through increasing excitement of experimental group children, or through "regression to the mean" from choosing either a low phase I interactor for the experimental group, or high phase I interactors for the control group. To avoid these and other problems, we adopt strict criteria for inclusion of data. The experimental child in each group is compared to the one other child closest to him in terms of phase I rate of volunteering; further, if no other child is within 2 units of the experimental child, data from that group are not included for analysis. This procedure controls for "excitment" effects (by comparing only phase I to phase III gain scores), controls for "regression" effects (by requiring phase I comparability between experimental and control group children), and substantially reduces the N upon which statistical tests are based (by using only one of the three control group children from every group). Use of these criteria resulted in exclusion of 2 of 20 groups of children.

References

Berger, J., and M. H. Fisek
1974 A generalization of the status characteristics and expectation states theory. Chapter 6 in J. Berger, T. S. Conner, and M. H. Fisek (Eds.), Expectation States Theory. Cambridge: Winthrop Publishers, Inc.

Brophy, J. E., and T. L. Good
1970 Teachers communication of differential expectations for children's classroom performance. Journal of Educational Psychology, 61 (October): 367-374.

Kervin, J.
1973 An information processing model for expectation states theory. Unpublished paper available from Department of Sociology, University of Toronto, Toronto, Ontario, Canada.

Entwisle, D. R., and Webster, M.
1972a Raising children's performance expectations. Social Science Research, 1 (June): 147-158.

1972b Raising expectations in the classroom. Unpublished paper presented at the annual meetings of the Eastern Sociological Society.

1972c Teacher expectancies and students expectation states. Unpublished paper presented at the annual meetings of the Pacific Sociological Association.

1972d Status factors in expectation raising. Unpublished paper presented at the annual meetings of the American Sociological Association.

1973 Research Notes: Status factors in expectation raising. Sociology of Education, 46 (Winter): 115-126.

1974a Expectations in mixed racial groups. Sociology of Education, in press.

1974b Raising children's expectations for their own performance: a classroom application. Chapter 7 in Expectation States Theory, J. Berger, T. Conner, and M. H. Fisek (Eds.). Cambridge: Winthrop Publishers.

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

Table 1

Mean Gains in Volunteering for Experimental and Control Children
Generalization Experiments

	Experimental	Control	S.E. of Difference	One-Sided t-value
White, rural Ss*	+1.14	+0.33	.544	2.04 ($p < .03$)
Black, urban Ss**	+1.10	-0.45	.614	2.52 ($p < .02$)

* 18 pairs

** 20 pairs

Table 2

Mean Gains in Volunteering for Experimental and Control Children; Direct Manipulation Experiment

	Experimental	Control
White, rural Ss*	+2.60	+0.57
Black, urban Ss*	+1.85	+0.97

* N = 20

EXPECTATION EFFECTS ON PERFORMANCE EVALUATIONS

Murray Webster, Jr.

Doris R. Entwisle

The Johns Hopkins University

Abstract

Situations in which expectations affect performance evaluations of actors are described and analyzed in terms of expectations states theory. Results of three experiments (N = 289) are presented testing derivations of that theory, and some implications of the results for theory building and for practical applications are discussed.

EXPECTATION EFFECTS ON PERFORMANCE EVALUATIONS[1]

Evaluations of performances, and by extension, of the actors making those performances, are important in a wide variety of social situations. Of particular interest to social psychologists are cases in which evaluations are influenced by factors other than objective criteria. A variety of processes--"halo effect" (Symonds, 1925), "cognitive set" (Kelman, 1961), group influence (Asch, 1956; Schachter, 1951), status effects (Caudill, 1958; Sherif et al., 1955), and others--have been proposed to account for discrepancies between objectively recorded performances and subjective evaluations of performance.

Distortion of evaluations in the direction of previous beliefs about abilities of the performing actors has been widely documented (see, for examples, Bales, 1970; Rist, 1970; Kelman, 1961). If one actor is thought to be good at a task, his performances are more likely to be highly evaluated than equal performances by an actor thought to have low task competence. One way to account for this type of cognitive distortion has been proposed by Berger, et al. (1972). They argue that problem solving interaction leads to formation of more or less enduring ability conceptions, called expectation states, for the actors involved. Once expectation states are formed and attached to actors, they are predicted to affect most important features of behavior: the higher the expectations held for a given actor, the more likely he is to be given and to accept chances to perform, to receive agreement and positive evaluations for any of his performances, and to exert influence over other group members.

Although this theory can account for many of the evaluation distortions previously mentioned, and in fact is claimed to apply to an even wider range of phenomena, direct tests and applications of the theory are still few. At present all direct tests have been conducted in laboratory settings, and every one of these tests has taken as its dependent variable rejection of influence in case of disagreement, only one of the many behaviors predicted to vary with expectation states. This limitation makes it hard to assess either the scope of the theory or its usefulness.

Knowledge of another's expectation state, or of a way to change it, would have enormous value in day-to-day living. Accordingly, the experiments reported here use a naturalistic setting and test how previous expectations affect future evaluations of performance. They also extend our previous work on the determinants and consequences of children's expectations for their own performance and how to change them (see Entwisle & Webster, 1972, 1974, for summary).

Expectation states theory forms a congenial context for analyzing interaction, since the schools may be seen as an arena where children perform and are evaluated. Teachers ask questions in class, give tests, and make homework assignments; students respond and their performance attempts are evaluated by the teacher and by other students. Moreover, evaluations of schoolchildren and their performances cannot be wholly based on objective criteria. For one thing, objective evaluative standards

are often lacking or ambiguous. What is a "good performance" in an esthetic judgment task? What, for instance, differentiates a good poem or a good painting from a bad one? Second, correct answers may be hard to recognize for difficult or complex tasks. Evaluating a single line of a computer program illustrates this point. In elementary school, peer evaluations, though very important to students, are voiced by peers who may not be capable of recognizing good or bad performances. Third, teachers often must evaluate students whom they have not had sufficient opportunity, or perhaps any opportunity, to test adequately. A mark in arithmetic may be based on inadequately sampled classroom behavior. Peer evaluations likewise are frequently based upon inadequate or biased test imformation. Fourth, often teachers are required to evaluate kinds of performance that neither they nor students can define. First grade teachers, for example, often give marks in "Language" or in "Spelling" when neither they nor their first graders can say exactly what "Language" or "Spelling" involves in first grade. Finally, expectations for some students are so firmly fixed on prior "evidence" that performance not in accord with expectations is disbelieved (a phenomenon also documented by Whyte, 1943, in the famous bowling incident of Street Corner Society, and lately for first-grade children by Seaver, 1973). Each of the above circumstances tends to confuse and attenuate evaluation solely by objective standards. To the extent that objective criteria are not used, expectations held for individuals making the performances will influence evaluations.

If objective evaluative criteria fail to specify evaluations exactly, then three possible situations, differing in degree of social definition provided by expectation information, are possible (see Berger et al., 1972): (1) In a maximally defined situation, performance expectations for the specific task at hand have been previously assigned to actors; for example, an actor may be known to possess high ability at task A, which are just the skills needed for successful task completion. Then it is a simple step to decide that any specific performance of that actor is probably a good one and should be highly evaluated. One expects a Metropolitan Opera singer to do well on a given aria even though one has never heard him sing it before. (2) In a less completely defined situation, ability is required at task A, but no expectations for performance of the actor at task A are known. However expectations are held for his performance at task B, and task B is known to be relevant to task A. The theory predicts that expectations will then be formed for task A which are the same as those already held for task B. If a person excells in snow skiing, we expect he will be good at water skiing. (3) In a minimally defined situation, ability at task A is required, ability at task B is known, but nothing is known about whether tasks A and B are related. In this case, so long as tasks are not explicitly dissociated (as they would be, for instance, by telling an actor that the skills are completely unrelated), the prediction is that individuals will conclude that the tasks are probably related. A research assistant who is good at computer programming will be expected to be good at planning laboratory experiments if nothing else is known about him. This process yields the same outcome as types (1) and (2) situations: the tasks are assumed to be relevant, expectations are assigned to actors, and these expectations affect the likelihood that actors will receive positive evaluations of performances.

One interesting consequence is that all three situations are predicted to yield identical outcomes for both expectation states and subsequent

behaviors of individuals. That is, so long as certain structural conditions are met, it does not matter whether the process is entered at stage (3) where only ability at task B is known, at stage (2) where in addition task B is known to be relevant to task A, or at stage (1) where ability at task A is also known. This consequence may be accurate but the only available data for direct test (Berger et al., 1972) do not provide particularly strong confirmation. Three experimental studies of the same assertions are provided here for social situations comparable to those of Berger et al. The concrete settings, the experimental design, and the measure of expectation states differ from those used previously, however.

Experimental conditions were designed to reflect the varying degrees of situational definition described above. Condition 1, which we call assignment, was maximally defined: subjects were told the ability of the other at certain tasks, this ability was made relevant to the particular task, and finally, expectations for the particular task were assigned to the other person. Condition 2, relevance, was less completely defined: subjects were told other's ability at another task and the ability was said to be relevant to the particular task, but no assignment of expectations to other was made. Condition 3, activation, was minimally defined: subjects were told only other's ability at another task, but no mention was made of possible relevance between tasks.

Experiments were conducted in the spring and replicated in the fall with different children in third grade classrooms of two suburban schools.[2] The investigator addressed the entire class, and told the children he was interested in finding out how well they could tell good words from bad words in sentences--the operational measure of "unit evaluation of a performance," which is predicted to vary directly with expectations held for other. (Good words were described as those which fit in well with the rest of the sentence, and which are exciting.) Then he handed each child a sheet with 10 sentences (Chart I). Each sentence contained some words ("performances") supposedly supplied by a (fictitious) other student. The fictitious students were described as having either high ability at schoolwork (Task B), or low ability. Instructions for the three conditions varied as follows.

> [all conditions] We have some words given by boys and girls with very high ability at most school subjects, and some by boys and girls with low ability. Students with high ability do better at reading, spelling, and arithmetic than students with low ability.
>
> [relevance and assignment only] Students with high ability at schoolwork also usually give better words than students with low ability.
>
> [assignment only] Today we have some words from students who have given good words in the past, and some from students who have given poor words in the past.
>
> [all conditions] However we have not yet graded the words you are about to see.

As mentioned, the set of experiments was run twice. Two classrooms of subjects were used for each condition on both occasions, making six classrooms per experiment, or twelve classrooms all told (the average of persons per classroom = 24.1). The sentences shown in Chart I were used in every classroom. In the first classroom of each condition, every odd numbered sentence was described as coming from a poor student. In the second classroom, descriptions of the fictitious authors were reversed: the even numbered sentences were supposedly given by good students.

Subjects graded each sentence by checking a box from the following:

<u>very good</u> <u>good</u> <u>fair</u> <u>poor</u> <u>very poor</u>.

Data were tallied by arbitrarily assigning "very good" a score of 1, and "very poor" a score of 5. Tables 1 and 2 present mean scores given to sentences, according to expectation condition.

It is evident from results in Tables 1 and 2 that in all three conditions expectations held for the author had marked effects on evaluations of his performances. In the majority of cases, when a given sentence was attributed to an author for whom high expectations were held, it was evaluated more highly than when the same sentence was supposedly produced by an author for whom low expectations were held. In Experiment I, this predicted effect of expectations on evaluations was observed 9 out of 10 times in the maximally defined (assignment) condition; 6 out of 10 times in the intermediate (relevance) condition; and 10 out of 10 times in the minimally defined (activation) condition. In Experiment II, the comparable proportions are 10/10 in assignment, 9/10 in relevance, and 10/10 in activation.

To assess the assertion that all three treatment conditions will produce equal effects on expectations, as well as to aid in further assessing the reliability of differences between expectation conditions, we present the results of an analysis of variance in Table 3. Fixed-effect factors are Treatment (high or low expectations for author), Condition (assignment, relevance, activation), and T x C interaction. Treatment and Condition both produce significant effects ($p < .05$), but the T x C interaction does not approach significance.

Discussion

These results show that, in three different conditions, expectations held for an individual affected evaluations of his performances. Although this effect is not surprising because it has frequently been pointed to in other settings, our work here apparently constitutes the first direct test of the assertion in an experiment excluding most other possible sources of variance. Results also indicate the considerable impact of peer expectations on evaluations given to other students. Peer evaluations are important factors in the classroom, so it is useful to confirm that the general relation between expectations and evaluations holds in this specific case. Both the laughter which greets certain students' answers,

and the awe which is accorded to others, seem interpretable in terms of expectation states theory.

Expectation states theory asserts that individuals will cognitively make the connections necessary to structure the incomplete "relevance" and "activation" conditions, and that as a result, these conditions will produce expectations identical to those given in the "assignment" condition. In a naturalistic setting such as the one we used, many factors which are uncontrolled can produce minor differences between conditions like ours, which all show statistically significant expectation effects. The most plausible alternative to the equality predicted by the theory is that conditions would be ordered by magnitude of expectation effect as follows: activation < relevance < assignment. There is no evidence in either set of experiments for this ordering from the analysis of variance shown in Table 3. Both Conditions (assignment, relevance, activation) and Treatments (high or low expectations) produce significant differences, but the Condition difference is the effect of classroom differences, not differences in information given to subjects. In order to establish any ordering of expectation effect, we would require a significant T x C interaction, and that clearly did not occur. In general, data in Tables 1 and 2 indicate that students in the two classrooms selected for the "relevance" treatment judged the words to be a little better than did the students in the other classrooms. The mean scores for all sentences in both expectation treatments were: assignment, 2.77; relevance, 2.43; activation, 2.69. This transformation of mean scores is what produced the significant main effect, not the experimental manipulation.

Either for theory development or practical analytic tasks, developing this new measure of expectations (evaluations of others' performances) should increase the flexibility accorded researchers. The measure used here is easy to implement and has straightforward relations with many cases of practical interest. Often it is not practical or possible to make use of disagreement resolution as a measure; as just one example, in many naturalistic situations there are norms which discourage the appearance or the recognition of disagreements themselves. In addition, the evaluations measured may be used in cases where the individual is not in direct interaction (such as disagreement) with the others for whom differential expectations are held.

Finally, we wish to note that, although our results show a tendency for individuals to base evaluations partially on data other than objective criteria, the tendency is not always harmful. To say that evaluations "should be" independent of expectation states seems as useless as saying that people "should be" unconcerned with status in their interactions. In many situations (such as those described earlier) individuals must reach evaluative conclusions when they simply do not have access to complete information from objective sources. In these situations, expectations held for actors--which result in most cases from observing previous performances of these actors--may well constitute the most adequate available basis for structuring an underdefined situation. What is rightly objected to is over-reliance on expectations in situations where objective criteria for evaluation are readily available.

Footnotes

[1]Experiments reported here were supported by Office of Education Grant No. 3-71-0122. We also acknowledge the research assistance of Muriel Berkeley, Beth Zaro, Michael Swafford, and Linda Olson with this work.

[2]We thank Dr. Evart Cornell and Mr. George Fanshaw, Principals, and the staff and students of Loch Raven and Stoneleigh Elementary Schools.

References

Asch, S. E.
1956 Studies of independence and conformity: a minority of one against a unanimous majority. Psychological Monograph, 70, no. 9.

Bales, R. F.
1970 Personality and Interpersonal Behavior. New York: Holt, Rinehart & Winston.

Berger, J., B. P. Cohen, & M. Zelditch.
1972 Status characteristics and social interaction. American Sociological Review, 37 (June): 241-255.

Caudill, W. K.
1958 The psychiatric hospital as a small society. Cambridge: Harvard University Press.

Entwisle, D. R., & M. Webster.
1972 Raising children's performance expectations. Social Science Research, 1 (June): 147-158.

1974 Raising children's expectations for their own performance: a classroom application. Chapter 7 in J. Berger, T. Conner, and M. H. Fisek (eds.), Expectation States Theory: A Theoretical Research Program. Cambridge: Winthrop Publishers.

Kelman, H. C.
1961 Processes of opinion change. Public Opinion Quarterly, 25: 57-78.

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

Schachter, S.
1951 Deviation, rejection, and communication. Journal of Abnormal and Social Psychology, 46: 190-207.

Seaver, W. B.
1973 Effects of naturally induced teacher expectancies. Journal of Personality and Social Psychology, 28 (December): 333-342.

Sherif, M., B. J. White, and O. J. Harvey
1955 Status in experimentally produced groups. American Journal of Sociology, 60: 370-379.

Symonds, P. M.
1925 Notes on rating. Journal of Applied Psychology, 7: 188-195.

Whyte, W. F.
1943 Street Corner Society. Chicago: The University of Chicago Press.

Chart I

Sentences Used for the Evaluation Task

1. There once was a very tall prince.
2. There once was a handsome movie star.
3. In order to fool the ladies, the man dressed up as a piece of furniture.
4. In order to fool the robbers, the princess dressed up as a big black bear.
5. The ocean was full of sharks and whales.
6. The parking lot was full of people with dogs.
7. When the Indians found the cowboy, they sat down and said hi!
8. When the teacher found the book, she opened it and began reading.
9. This was a good thing to do on Sunday.
10. This was more than he wanted.

Table 1

Experiment I: Mean Scores of Sentences by Expectations for Author

Sentence	Assignment High	Assignment Low	Assignment Difference*	Relevance High	Relevance Low	Relevance Difference	Activation High	Activation Low	Activation Difference
1	2.20	2.70	+ .50	2.00	2.39	+ .39	2.00	2.50	+ .50
2	1.42	2.76	+1.34	1.81	2.31	+ .50	1.68	2.04	+ .36
3	3.25	3.80	+ .55	2.76	2.84	+ .08	3.18	3.82	+ .64
4	2.46	2.64	+ .18	2.13	2.56	+ .43	2.68	2.86	+ .18
5	1.83	2.68	+ .85	1.88	1.94	+ .06	1.43	2.64	+1.21
6	2.96	3.79	+ .83	2.84	2.82	- .02	2.46	2.78	+ .32
7	4.27	3.64	- .63	2.12	3.65	+1.53	3.29	4.07	+ .78
8	1.92	2.77	+ .85	2.00	1.82	- .18	2.04	2.44	+ .40
9	2.33	2.48	+ .15	2.47	2.41	- .06	2.37	3.00	+ .63
10	1.87	2.67	+ .80	2.41	2.35	- .06	2.14	2.86	+ .72
Mean	2.45	3.00	+ .55	2.24	2.51	+ .27	2.32	2.89	+ .57
(N)	25, 30			17, 32			28, 28		

* positive difference indicates predicted direction.

Table 2

Experiment II: Mean Scores of Sentences

by Expectations for Author

Sentence	Assignment			Relevance			Activation		
	High	Low	Difference	High	Low	Difference	High	Low	Difference
1	2.04	2.79	+ .75	2.20	2.45	+ .25	2.05	2.26	+ .21
2	1.71	2.44	+ .73	1.64	2.52	+ .88	2.21	2.90	+ .69
3	2.52	3.79	+1.27	2.54	4.18	+1.64	2.68	4.47	+1.79
4	1.89	2.70	+ .81	1.91	2.44	+ .53	2.26	2.79	+ .53
5	1.38	2.57	+1.19	1.96	2.91	+ .95	2.53	3.24	+ .71
6	2.39	3.31	+ .92	2.82	3.32	+ .50	3.06	3.79	+ .73
7	3.11	4.39	+1.28	2.88	3.64	+ .76	3.28	3.89	+ .61
8	1.75	2.41	+ .66	1.55	2.12	+ .57	1.74	2.37	+ .63
9	2.18	3.21	+1.03	2.60	2.00	- .60	2.05	3.26	+1.21
10	2.25	2.56	+ .31	1.36	2.68	+1.32	2.11	3.47	+1.36
Mean	2.12	3.02	+ .90	2.33	2.90	+ .57	2.40	3.24	+ .84
(N)	27, 28			25, 11			19, 19		

Table 3

ANOVA of Overall Mean Scores

Source of Variation	df	Mean Square	F	p
Treatment	2	1.16	3.30	< .05
Condition	1	12.07	34.39	< .05
Replication	1	0.08	--	n.s.
T X C	2	0.20	--	n.s.
Within T X C X R	113	0.351		

Psychological Reports, 1972, 31, 551-556.

EFFECT OF A PRINCIPAL'S EXPECTATIONS ON TEST PERFORMANCE OF ELEMENTARY-SCHOOL CHILDREN[1]

DORIS R. ENTWISLE,[2] EVART CORNELL,[3] JOYCE EPSTEIN[2,4]

Johns Hopkins University

Summary.—A 17-classroom experiment shows that a (male) principal's comments conveying positive expectations to boys on a first test are followed by improved performance on a second test. No effect of comments was apparent for girls. The experiment is discussed in the context of current research on teacher expectancy.

Expectations of teachers for students and how these expectations affect performance have been the topic of much recent research (Brophy & Good, 1970; Claiborn, 1969; Fleming & Antonnen, 1970; Meichenbaum, Bowers, & Ross, 1969; Rosenthal & Jacobson, 1968) and also of some controversy. In most of the studies where teachers' expectations are presumed to influence pupils, no actual measurements of teachers' behavior or attitudes are made. One exception is work by Brophy and Good (1970) who in studying dyadic contacts between teachers and students observe that teachers demand better performance from those students for whom they hold high expectations. Another exception is the study of Rothbart, Dalfen, and Barrett (1971) who in a simulated classroom situation show that teachers are more attentive to students for whom they hold high expectations. More work is needed on how school personnel convey their expectations and how this shapes students' behavior. The present report of a school-wide experiment is an attempt to provide such information on how expectancies are mediated.

A study of teacher comments and student performance by Page (1958) suggested that on one objective test teachers' comments had a positive effect on a student's ranking on a subsequent objective test. In other words, those students receiving comments on a first test ranked higher than other students receiving no comment when a second test was given. Written comments can be considered as one mediator of expectations.

Because of the current interest in expectancy and performance, we decided to pattern an experiment after Page's but to use comments conveying *only* expectations (see below) and to explore the range of "significant others." Our major interest lay in whether comments by the *principal* on arithmetic test papers

[1]This research was supported by Office of Education, OEG Grant No. OEG-3-71-0122. The opinions expressed in this publication do not necessarily represent the position or policy of the Office of Education, and no official endorsement by the Office of Education should be inferred.

[2]Department of Social Relations, Johns Hopkins University.

[3]Principal, Stoneleigh Elementary School, Baltimore County, Maryland.

[4]We wish to thank the teachers in Stoneleigh School for their help with this research. We also wish to thank Dr. James Fennessy for many helpful suggestions.

of elementary school students stating that he had positive expectations would lead to improved performance on the next test. The question at issue is whether expectations of a school principal conveyed in a definite but rather distant manner will lead to improvements in performances. Is the school principal, as well as the teacher, a "significant other" in students' eyes?

EXPERIMENT

Teachers in a white middle-class elementary school in the suburbs of Baltimore gave a routine test in arithmetic. A mark was assigned the paper by the teacher (A, B, C, D, E). The papers were then collected and given to the principal, who had announced to each class before the first test (17 classes, Grades 2 through 6) that he was going "to look over the papers" and that he would be "writing comments on some of the papers." For each letter-grade stratum a random division determined those papers to receive comments and those to receive no comment. Standardized comments (see the list below) calibrated to the mark the paper was given were placed on a random half of the papers in each class.

Grade Given Paper by Teacher	Comment Added by Principal
A	I know you will do as well next time.
B	Maybe next time you can do even better.
C	I feel you can do even better than this.
D	I know you can do better than this on the next test.
E	I am sure you can get a better grade on the next test.

The comments were phrased to emphasize *expectations*. Comments directly conveying reinforcement (praise) were deliberately avoided. Within a few days the papers were returned to the students and shortly thereafter a second arithmetic test was administered by the classroom teachers. Scores on this second test are the criterion. The two classroom tests were tests that teachers would be giving routinely. That is, in all grades in this school every week or two, arithmetic tests are routinely given to check pupils' progress and to provide feedback for the teacher and students. The experiment was superimposed on this ordinary test-giving pattern. Some teachers gave tests 1 wk. apart, others 2 wk. apart. The tests covered material that was being taught during the time of the experiment and this of course differed from grade to grade and also from teacher to teacher. This kind of testing, since it is part of standard teaching practice, provides a suitable vehicle for checking effects of comments. If we had tried instead to alter normal teaching practices any effect of comments would be confounded with other alterations introduced.

Prior to the experiment, teachers were asked to assign a "suggestibility" rating to students in their class on a 5-point scale as follows: (1) "highly re-

sponsive;" (2) "moderately responsive;" (3) "no obvious response;" (4) "highly reactive" (negative); and (5) "highly reactive" (positive). This variable was included on the supposition that some children might be unresponsive to comments and therefore that the experimental treatment would be ineffective for them.

Test scores were transformed from letter grades to numerical values (A = +2, B = +1, C = 0, D = −1, E = −2). The zero-order correlation matrix presented in Table 1 shows that variable 7 (Treatment) does not correlate significantly with variable 9 (2nd test score) when boys' and girls' scores are grouped together. However, the correlation between treatment condition and posttest score for boys is .16 (n = 217), significant at the .02 level, and for girls is —.05 (n = 225), not significant. (The correlations between pretest scores and treatment are only .02 and —.05 for boys and girls respectively.) For the 17 classes, if test gains are compared by sex for the comment and no-comment conditions there is an average gain of 2 units by class for boys and a decline of almost 1 unit by class for girls. A t test of this sex difference shows that it is significant beyond the .05 level (t_{16} = 2.21, classes as the unit of analysis). The difference between test scores of individual boys in the comment and no-comment groups is about half a letter grade. This difference of half a letter grade for boys brought about by the experimental treatment is not large in absolute terms. On the other hand, it followed a "small" treatment, the writing of a single comment on one classroom test.

The experimental treatment did not interact with the students' rated responsiveness. That is, those students rated by teachers as "highly responsive" or "moderately responsive" did not appear to be any more affected by a principal's comment than students rated as "not responsive." As one would expect,

TABLE 1

ZERO-ORDER CORRELATIONS BETWEEN VARIABLES, GRADES 2 TO 6 (N = 442), M_{IQ} = 115

Variable	1	2	3	4	5	6	7	8	9
1. IQ*		.28	.19	.58				.18	.20
2. 1st semester math grade*			.66	.52	.27			.47	.47
3. Effort Grade*				.27	.45			.32	.37
4. Math Achievement*								.36	.37
5. Suggestibility**						.14		.24	.30
6. Sex									
7. Treatment (comment vs no comment)†									
8. 1st Test (transformed)‡									.49
9. 2nd Test (transformed‡									

Note.—Only correlations significant at or beyond the .01 level are listed.
*From school records. **Teachers' judgments. †Experimental treatment. ‡Grades assigned by teachers.

responsiveness as judged by teachers correlates with sex (girls being more responsive) and with all teacher-assigned grades (first semester mathematics grade, effort grade, and grades on classroom tests) but not with scores on standardized tests (IQ and mathematics achievement).

Discussion

The chief finding is that boys' scores on a second arithmetic test improve when the boys received a comment from the principal conveying positive expectations on a first test, relative to a control group. Girls' performance on second test was not affected by his comments on the first test.

The significant correlation between experimental treatment and posttest score for boys, associated with an average increase of half a letter grade, must be weighed against several factors, first the "smallness" of the treatment already mentioned. Second, arithmetic is a subject where trying harder may have little effect. In some areas like language arts, where trying harder might lead to longer or more elaborate responses, comments might yield more striking immediate effects. Third, the correlation between treatment and boys' posttest grades is much like the correlation between standardized tests (IQ and math achievement) and posttest grades. Unreliability of classroom tests attenuates relationships between these tests and other variables including experimental treatment.

The present study is consistent with earlier work. Page reported a significant difference in favor of comments, but his experiment differs in several important ways from the present one—in particular there is no breakdown in terms of sex so one cannot tell whether there was a differential effect for boys and girls, and comments were delivered by teachers rather than by the principal. It is hard to tell how large Page's comment-effect is. Apparently, triplets of students in the 3 experimental conditions (personalized comment, standardized comment, no comment) were ranked 1, 2, 3 according to performance on the first test and then re-ranked on the basis of the scores on the second test. For 74 classes with about 30 per class, the largest difference in mean rank between means of treatment groups is 0.18. This could stem from small differences in absolute performance level although no precise statements can be made from Page's published results. When our data are ranked in an analogous way, it appears that the magnitude of the effect in Page's experiment and the effect in the present study may not be far different.

Two plausible mechanisms could explain the differential effect of comments by sex. (1) As the principal delivering the comments was male, he might be persuasive for boys since most other school personnel were female. (2) In other research in the same school where expectations of children for their own performance were raised in a short experiment, boys were generally found to be more responsive to the experimental expectation raising treatment than girls

(Entwisle & Webster, 1972). This treatment involved an individual session with an adult where all responses in a simple story-generating task were heavily reinforced. It was thought that boys might be relatively deprived of praise from teachers—several workers report that boys receive less praise and more blame from teachers than girls (Felsenthal, 1971; Jackson & Lahaderne, 1967; Meyer & Thompson, 1956) so that their expectations initially are at a generally lower level and therefore easier to raise. The same line of reasoning can be applied here to explain the greater effectiveness of the principal's comments for boys—because they have received few positive comments they may be especially receptive.

This research is related to an expanding body of findings on teachers' expectancy (e.g., Meichenbaum, Bowers, & Ross, 1969) and suggests why some results from research on teachers' expectancy could be equivocal or negative. [We intentionally avoid raising issues about the methodological flaws others see in this research, for example, Thorndike (1969).] First, teachers' expectancies represent expectations of one set of significant others, but expectations of peers, of parents, and as shown here, of other school personnel, are probably operating concomitantly. If expectations of various "others" are contradictory, they could work against one another to attenuate effects in studies of teachers' expectancies. Second, research has so far paid little attention to the status characteristics of the person whose expectations are altered. A middle-aged female teacher may have low credibility for a 10-yr.-old boy, so that her increased expectations for him are relatively impotent in affecting his behavior. Much further work needs to be done on exactly how expectations are transmitted. The present study is a small step in that direction.

To sum up: expectancy comments by a school principal (male) apparently are effective in improving boys' performance. The size of the increase, although statistically significant, is small in absolute terms. Whether such an effect could be produced repeatedly for the same students and whether the effect would persist should be studied.

REFERENCES

BROPHY, J. E., & GOOD, T. L. Teachers' communication of differential expectations for children's classroom performance. *Journal of Educational Psychology*, 1970, 61, 367-374.

CLAIBORN, W. L. Expectancy effects in the classroom: failure to replicate. *Journal of Educational Psychology*, 1969, 60, 377-383.

ENTWISLE, D. R., & WEBSTER, M. Research note: status factors in expectation raising. *Sociology of Education*, 1972, in press.

FELSENTHAL, H. Pupil sex as a variable in teacher perceprion of classroom behavior. Unpublished paper presented at February, 1971, meetings of American Educational Research Association, New York, New York.

FLEMING, E. S., & ANTONNEN, R. G. Teacher expectancy or my fair lady. Unpublished paper presented at the annual meetings of the American Educational Research Association, 1970.

JACKSON, P., & LAHADERNE, T. Inequalities of teacher-pupil contacts. *Psychology in the Schools*, 1967, 4, 204-211.

MEICHENBAUM, D. H., BOWERS, K. S., & ROSS, R. R. A behavioral analysis of teacher's expectancy effect. *Journal of Personality and Social Psychology*, 1969, 13, 306-316.

MEYER, W., & THOMPSON, G. Sex differences in the distribution of teacher approval and disapproval among sixth-grade children. *Journal of Educational Psychology*, 1956, 47, 385-396.

PAGE, E. G. Teacher comments and student performance: seventy-four-classroom experiment in school motivation. *Journal of Educational Psychology*, 1958, 49, 173-181.

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

ROTHBART, M., DALFEN, S., & BARRETT, R. Effects of teacher's expectancy on student-teacher interaction. *Journal of Educational Psychology*, 1971, 62, 49-54.

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

Accepted July 5, 1972.

PROCEDURES THAT TURNED OUT TO BE UNSUCCESSFUL

Murray Webster, Jr.

Procedures That Turned Out to be Unsuccessful

We report here three different procedures which we found to be unsatisfactory for the study of classroom expectations, or for intervention in the natural course of development of expectations. Each procedure was attempted in the course of a particular phase of our investigations; that is, it was developed for a purpose related to the conceptualization presented in the body of Section I. We describe them here sequentially, as they entered the course of the research program.

A. The Artistic Judgment Task

From the beginning of the experimental program, we have been concerned with studying the sources and the consequences of expectations in general, rather than studying expectations for some particular skill such as reading, or story telling ability. As we noted, because of the existence of the story telling task developed in previous research, we first sought to adapt this task for our expectation raising experiments. However immediately after the first successful attempts to raise expectations with this procedure, we began to search for other tasks which might be used for the same purpose. Our thinking was that the early demonstrations of our technique would be more persuasive if they could be shown to apply to more tasks than the story telling. However as we noted earlier, because of the stringent set of requirements which a task must meet for this work, task development is not an easy problem.

One task which seemed promising we called "Artistic Judgment Ability." Artistic Judgment Ability met all the requirements for our experiments--such as being interesting to children, consisting of discrete performances, permitting either positive or negative evaluations of any given performance--and it offered the additional virtue of being quite different from the story telling task. Thus we could argue if the experiments with it worked that the procedure for raising expectations was probably a very general one, not tied to a particular sort of task.

During the fall of 1971 we spent considerable time in pilot studies of Artistic Judgment Ability. The basic experimental design for raising expectations already described was used with only the task differing. Artistic Judgment Ability was presented to children as being the ability to discriminate artistic merit in drawings; to tell good drawings from bad ones, in other words. Children were shown a series of figures like Figure 1 and were asked to "grade" them from A to E. This presentation was done in the classroom, using booklets of 10 drawings. Booklets were collected, suposedly to be graded by the experimenters, and after a time

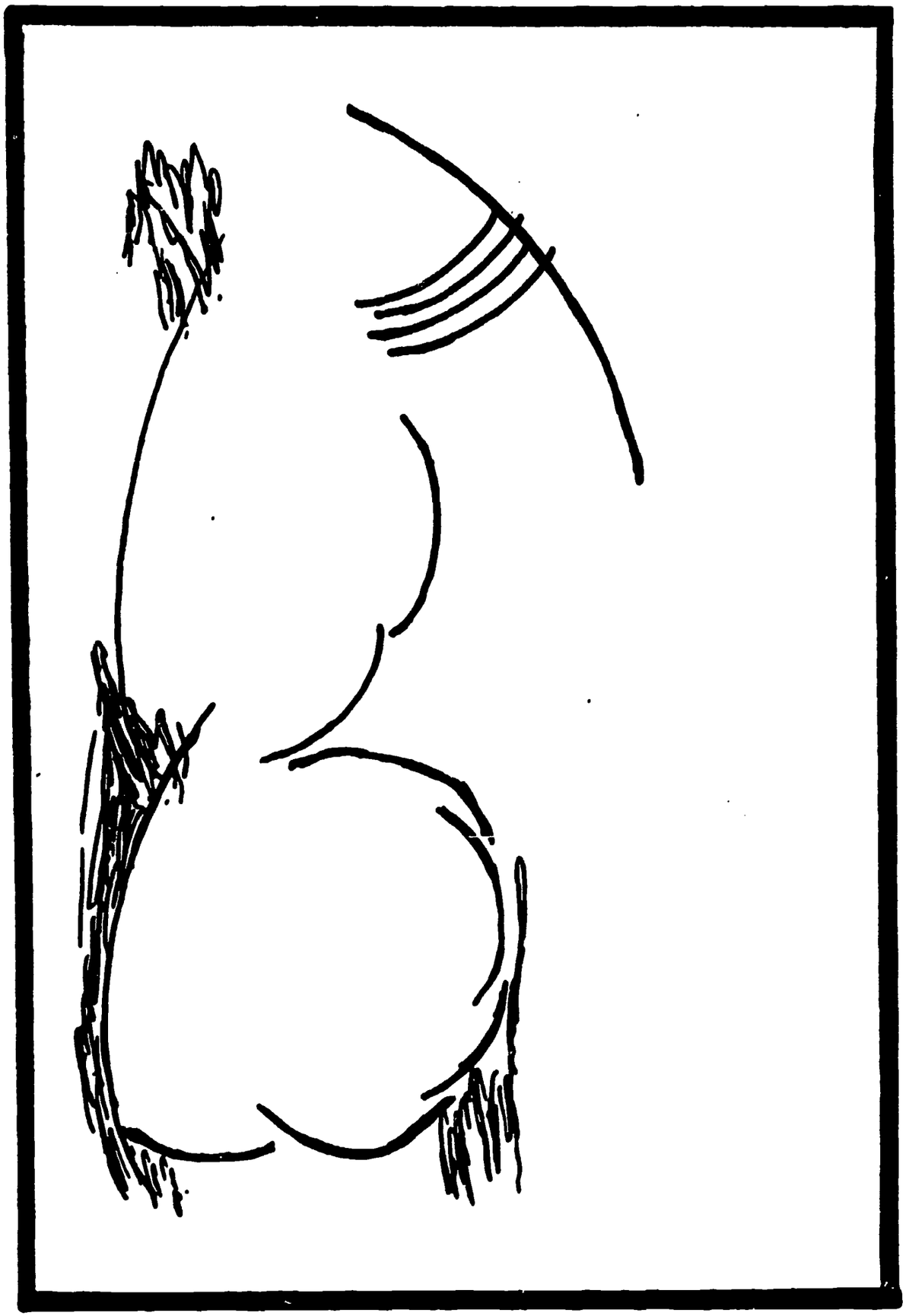


Figure 1

were returned to the children. They were also at that time told that their Artistic Judgment Ability was either "very good," or "average."

Later that same day, groups of four children were assembled, containing one "experimental" child who had been told she was very good, and three "control" children who had received scores of average. Subjects for these tests were white middle-class girls, third and fourth graders. In each group, 12 new drawings were shown to the group, and children asked to volunteer to grade them.

Structurally, this experiment is the same as our basic experiment: First, the attempt is made to induce high expectations in one randomly selected (experimental) child. Then the experimental child is put into a group with three children of presumably average ability, they all are given a set of 12 action opportunities, and are given the chance to emit performance outputs. The performance outputs, or problem solving attempts, are then recorded for each child and used as the measure of expectations held. The prediction is that the child with high induced expectations will be more likely than control group children to accept a given action opportunity and attempt to perform.

Results were essentially zero: our analyses show no effect of the earlier manipulation procedure. We ran a few groups to see whether the time lag between manipulation and measurement phases could be responsible--children might have forgotten their supposed ability levels--but that did not seem to be the case. Also, interviews with children usually elicited strong evidence of remembering what ability we had assigned to them.

Of several possible reasons for the failure of this task, the one which seems most reasonable to us is that "artistic judgment" is socially defined to be a skill difficult to measure and without universalistic standards. One person's artistic opinion is considered as good as the next person's, that is, there are no clear distinctions between "good" and "bad" when it comes to art. Some children volunteered ideas similar to this one, though in most cases they simply said they saw no reason to volunteer more often than anyone else in the group setting.

What this interpretation indicates, of course, is that Artistic Judgment fails to satisfy the major scope condition for Expectation States Theory, namely, task-orientation. Artistic Judgment seems to belong to the realm of process orientation (in Bales' terms), not susceptible to clear or objective assessment. The meal planning task which we later developed and used successfully meets the same sorts of criteria which prompted us to develop Artistic Judgment, and it turned out to be much more successful.

B. The Allocating Action Opportunities Experiment

During the summer, 1972, we began to get interested in effects of peer expectations and behavior on the expectations children develop for themselves. On an intuitive basis, it seemed clear to us that children do

develop ideas of how good each other is at various school tasks, and they show these ideas in a variety of ways. One important behavioral consequence of expectation states held for others is predicted to be differential allocation of action opportunities: all other things being equal, the theory predicts that individuals will give chances to perform more frequently to those others for whom they hold high expectations than to those for whom they hold low expectations. Also, our theoretical analyses of classroom interaction indicates that the differential allocation by the teacher of chances to perform is one of the main ways in which he or she conveys the expectations held for individual students. (Such a process operates, we believe, in the famous "Pygmalion effect.") The allocation of action opportunities experiment was designed to deomonstrate this effect.

The experiment devised was a modification of our basic expectation-raising experiment. Groups of three children were assembled for the story telling task, and one additional child--whose later behavior was to constitute the dependent variable measured--observed the group make up a story. Phase I of this experiment differed from the basic design in that every performance given in Phase I was evaluated publicly by the experimenter. One of the three children received consistently positive evaluations; the other two received about half positive and half negative evaluations. (These performances were not volunteered; each child was selected in turn.) The intent of this procedure was to raise the "observer" child's expectations for one of the three children relative to the other two. In Phase II, the four of them were defined as a "team," in competition with other teams, and with the former observer as Captain. As each sentence of the story telling task was read, the Captain would decide which of the other three children would give the team's word. There was no opportunity to volunteer. The prediction, of course, was that more action opportunities would be given for whom higher expectations were held.

What happened was that the Captain invariably distributed action opportunities on a strictly equal basis. Sometimes he called on each child in rotation, sometimes alternating, but always with striking equality. At first we thought this might be due to lack of task orientation, so we increased the competition between teams by giving prizes and by making the other teams visible to the children before and after Phase II. We also thought the number of children might be too large for remembering expectations, so we reduced the "performing" group to two members. We also ran some four-person groups. We investigated the effects of natural status by making the Captain older than the performers (fourth grader with second graders), and by using children all of the same age. None of these variants in the social composition or the structure of the rules produced any deviation from strictly equal distribution of action opportunities.

We conclude that the failure of this design was due to operation of very strict norms of "parity," "fairness," or "equity," as they are variously called. Basically, in our culture, there are rules that everyone on a team should get a chance to play, and there are rules in the school system which discourage making differential evaluations of individuals. (It goes without saying that these are hypocritical norms,

publicly espoused but seldom actually followed.) For our Captain to allocate action opportunities differentially, he would have had to violate these norms, and apparently this was not possible for him. Incidentally, during the course of our attempts to develop a successful allocation experiment, we learned of problems Joseph Berger had had at Stanford in the social psychological laboratory where he was attempting to devise an allocation experiment. Even with more control in that setting, and with a college age population, no satisfactory means was found.

The final set of our classroom experiments (in which children "grade" stories supposedly made up by other children) deals with a closely related issue, and it was successful. A second consequence of expectations states held for others is differential allocation of performance evaluations, and although here too we had problems with the "fairness norms," this procedure eventually was quite successful in producing the predicted results. To our knowledge, however, there exists at this time no successful demonstration of the theory's predictions regarding allocation of action opportunities.

C. Measurement of Self-Esteem

For the longitudinal study of the natural course of expectation development (described fully in Section II of this report), it was desirable to develop ways to measure expectations children held for themselves and others at the tasks they actually perform in the classroom. The educational literature contains numerous measurement techniques--typically, these are developed by a single researcher, not particularly well validated, and never appear again. However because of the obvious advantages of adopting a standardized measurement, we tried out several of these for our own work.

Beginning before this project was undertaken (July, 1971), we worked with a procedure to assess sociometric ranking of children by asking them to choose each other for hypothetical teams which they were told would then compete for some prize. Every child in the classroom is asked to put on a paper hat. The hats each have three colors; for example, red at top, blue next, and green brims. However the order of colors may differ, so that, for example, a third of the children have hats with red at the top, a third with green at the top, and so on.

Children are lined up so that they can see everyone else in the room, and they are asked repeatedly to write their first three or four choices in order--choosing first from those with red at the top, next from those with green at the top, etc. Some previous literature has also reported putting numbered signs around each child's neck, and sometimes in addition numbers of hats are changed during the procedure.

Mathematically, it is possible, with a small number of repetitions of this procedure, to determine the entire sociometric ranking of a large class of children. Operationally, it was impossible. Children quickly became bored with the procedure, they changed hats without our knowledge,

they didn't follow instructions about which group to rank from, and in general both the actual operation and the results were too confused to be useful. With the wisdom of hindsight, we are not surprised by this outcome; we are, however, amazed that other investigators report successful use of this technique.

Another technique with some precedent in the literature is the Draw A Man task. Children are asked to draw or copy a series of stick figures, and their drawings then are scored according to presence or absence of certain graphic characteristics (such as elaborated vs. simple, with hands vs. no hands, etc.). The technique is intended to operate similarly to other projective measures such as the TAT, and some investigators have reported success using it to measure self-esteem.

After some attempts to use the Draw A Man task with children of different ages (first thought fifth grades), we concluded that it was not useful for that purpose. The most important factor in this conclusion was extremely low test-retest reliability and inter-item reliability. However in addition, children's overall self-esteem scores on this measure did not correlate well with anything else known about them, such as grades, popularity as judged by teachers, or other measures of expectations.

At this point our view of measures of self-esteem is that even if a measure with good reliability could be found and applied to different social groups, it would not be useful for the sort of theoretical and empirical work we are engaged in. The concept "self-esteem" is given a wide variety of meanings in the social psychological literature, and there usually is little overlap between what different researchers mean by it. What is notably lacking from all the self-esteem literature of which we are aware is the two properties which are central in our conception of expectation states; namely, expectations are relative to a particular task rather than general ideas of competence, and they are relative to a particular social setting rather than being fixed ideas an individual carries around with him wherever he goes. Lack of specified tasks and specified comparison others has produced some very strange findings in previous work, as we noted in Section I of this report.

The procedure we eventually developed, where a child or a parent is asked to estimate the child's future report card in reading, arithmetic, and conduct, is relational in both the senses above. In addition, it has been shown to produce high test-retest reliability, and expectations of parents and children as measured by this procedure are predictive of a large number of other behaviors and performances, as we show in Section II.

SECTION II

OBSERVATIONAL STUDIES OF CHILDREN'S EXPECTATIONS

Doris R. Entwisle and Leslie A. Hayduk

CHAPTER 1
INTRODUCTION

Sociologists pay little attention to persons below the age of adolescence. Modern society, nonetheless, thrusts children outside the family early when children become members of day-care, nursery, or kindergarten groups. These groups are structured, with many of the characteristics of adult work groups. Most children commute daily to their groups, eat at least one meal with their groups, obtain various medical and health services under group auspices, and find considerable entertainment and recreation, as well as "work," with their group.

Nothing much is known about a child's early experiences in group life and group activities--his sociological upbringing, if you will. How does this upbringing affect his later functioning as a group member or affect the nature of adult social groups? Theorists like Mead and Cooley who talked of the "looking-glass self" supposed that the developing person shapes his self-image, especially his ideas about his own ability, from reactions others display when he performs or interacts with them. A young child's social matrix outside the family provides the first looking-glass. But a detailed account of how social process shapes the child and how a child in turn shapes social process remains to be given.

In the first part of this report we have presented experimental studies which show explicitly how some social processes affect children's expectations. (Expectations can be roughly defined as a child's concept of his own ability in a particular activity.) In small groups of four children our experiments show that a child's expectations for his own performance can be raised by a suitable adult. Related experiments show that an improvement in expectations for one task spills over to improve expectations for other tasks even when the two tasks seem unrelated. We also show that children themselves shade their judgements of the quality of another child's performance depending on whether they believe that child has previously performed well or poorly. Exactly the same performance, in other words, gets a different evaluation depending upon the identity of the person emitting the performance. Others (Finn, 1972; Seaver, 1973) have shown a similar shading of judgements by teachers when they possess prior information which gives rise to expectations. A child's early groups, then, to the extent they resemble our experiments, could build a child's self-image to be consistent or inconsistent with his ability.

Our experiments aimed at specifying exactly how expectations may be changed and by whom. They are, in short, an attempt to specify in fine grain how the social "looking-glass" operates to lead a child to form ideas about himself and his abilities, and what the consequences are in terms of the child's performance.

A noteworthy fact which should not be overlooked although we have paid no attention to it in our experimental work, is that children's expectations are not null at the start of our experiments. Children come to our experiments with expectations based on whatever has happened to them in the past. Each has a life history filled with events that forecast success or failure for the future, that set the level of expectations for success at all kinds of activities the child has not yet engaged in, and that cause the child to choose to perform or not to perform. Some children come to our experiments with expectation of doing well even though they have never done anything exactly like this before. Other children come with pessimistic views about doing well. The two act very differently. How do they get this way?

This, the second part of our report, attempts to trace out early life events that may shape expectations. Observational findings will be presented. The purpose of the observations is to trace individual children, starting in first grade and sometimes a little before, to see what happens to their expectations as they go through the early years of school. When children are in first grade how do they think they will do on their first report card? After the first report card, do they modify their expectations? Are parents' and peers' expectations inportant in shaping the child's expectations for himself, and if so, how important? Do expectations for children in mixed racial groups follow a different pattern or the same pattern as in segregated groups?

What Are Expectations?

Expectations for the self are a set of beliefs loaded with affect which may or may not correspond with reality. They provide one of the chief social linking mechanisms between members of a social group. As Hamburg (1963) says:

> Man is by nature committed to social existence...man's self-interest is best served through his commitment to his fellows...[;] [his] need for positive affect...may be expressed...as a desire for...recognition and acceptance, for approval, for esteem, or for making persons...act in such ways as to attain the approval of their fellow men.

Each persons craves response from other humans and, as others have often noted, there is an insatiable need for these responses to maintain and enhance the phenomenal self. Expectations for the self, then, are the set of beliefs an individual holds about himself as a consequence of social feedback from others, and they span the whole set of attributes that contribute to self-regard.

When the child emerges from the protective circle of the family he begins to get social feedback from new sources--teachers, other school personnel, or a wider set of playmates and peers in school--as a response to new sorts of activities. He is responded to and evaluated on the basis of his own achievements, things like his ability to please the

teacher, to impress his age-mates at ball playing or running, to wear clothing other children admire, to forecast others' responses to him, and, very much, to achieve in those academic areas like reading that the culture values highly. The standards of judgment may be very unclear to the child (as our data indicate) but he is not in doubt about whether others are according him esteem or barely tolerating him. He internalizes these evaluations and uses them to structure his expectations for himself.

Once the child enters school the process of social comparison begins in earnest. His expectations for himself can be the bulwark supporting the acquisition of competent behaviors if evaluations from others are supportive. Favorable other-evaluations are translated into favorable self-evaluations. As will be seen, our data suggest that most children have very optimistic expectations for themselves when they start school--the vast majority (80% or so) expect to get the highest possible mark in reading. Obviously, however, the process of social comparison becomes more discriminating as the social reference group enlarges. In the family what little social comparison there is on the basis of achievement is mitigated by ascriptive characteristics (the 4-year-old is not expected to perform as well as his older sibling) or even distorted by ascriptive characteristics (the pretty child may be more warmly regarded than the ugly child even though both perform to the same objective standards of excellence in behavior).

In the school ascriptive characteristics are less important because the arena of behaviors is greatly enlarged by the intrusion of achievement behaviors, reading, arithmetic, and the like, so the basis for social comparisons differs. More important, the numbers of individuals involved in the social comparison process greatly expands so that feedback now comes from many more sources and evaluations are made with respect to many more fellows. Only one child can be "best" in reading so with 30 children in a class, 29 suffer some loss of reward. There is some experimental evidence (Gibby and Gibby, 1967) that failure causes a drop in the self concept and a perception of being less well-regarded by significant others. If failure is less than complete, i.e. if someone else outdoes a child in a given activity, the child may still perceive himself to fail, as he does in a relative sense. It is little consolation for a student who gets a C in 4th-year French to realize that he can understand French better than 99% of Americans. He is concerned with his performance relative to other 4th-year French students. Similarly the child in a middle-class school who receives a low mark in reading in first grade will think of himself only in terms of others in his class and their marks.

There is also the potential, with more evalutators, for evaluation to differ. A little evidence (Kerensky, 1966) indicates that pupils' self-perceptions can differ significantly from teachers' perception of them and that the two have different sets of expectances regarding school behavior. As will be seen, we find that average expectations of parents and teachers are equivalent but that child-by-child the parent is only a little better than chance at predicting how the teacher will evaluate his

child. It is very easy, therefore, for the parent to expect an A while the teacher expects, and awards, a C.

In any case the early days of school plunge the child into a new environment of social comparison and the net residue of these comparisons shapes his evaluation of himself, what we call his "expectations." His views of himself are thought to be a crucial component in complex mental processes, filters if you will, which color and even forecast his experiences. Lamy (1965), for example, found for 52 first-grade students in a laboratory school at the University of Florida that self perception in kindergarten predicted reading achievement in first grade as well as IQ, and Wattenberg and Clifford (1964) could predict reading achievement two and one half years later from measures of self-concept procured in kindergarten. The later study indicates, besides the importance of self-concept for reading, the causal priority of it.

Why Study Expectations?

Several large-scale studies make the consequences of holding high or low expectations abundantly clear. In the famous Coleman (1966) Report, for example, minority group members revealed that they had low expectations for their own effectiveness in controlling events in their own lives and it turned out the lower their feelings of control, the poorer their academic records. There was no direct evidence available, however, as to how and when these ideas were acquired. Also in Brookover's (1962, 1964, 1965, 1967) longitudinal study of students from the seventh through the twelfth grades, students' self-conceptions of ability were shown to predict performance better than IQ. We do not know how these self-conceptions of ability were initially established, or what led some students to have high self-concepts and others low. (Shaw, et al., 1960, also does a correlational study of self-concept and academic achievement.)

Surprisingly, there are no studies which focus on academic self-concepts of younger children or on how academic self-concepts are established in the first place. Lesser (1972) says there must be important effects when "the child exhibits his elementary skills like naming letters or numbers in the presence of someone who cares about him and receives attention and admiration." But how in fact a child develops as image of himself as a competent and effective, <u>or</u> incompetent and ineffective person is shrouded in mystery.

Young children are much harder to study than older ones, and this in large part is responsible for the dearth of research upon expectations and aspirations of children in their early school years. What is to be gained from this added effort? The answer is "much" on two counts. First, folk wisdom and a wide array of scientific evidence suggests that by the end of third grade a child has developed a fairly complex and stable self-image. Performance levels at that time are surprisingly good predictors of subsequent performance. In Husen's (1969) large cross-national study, for example, intelligence scores and teachers' ratings in third grade were very good predictors of subsequent educational

careers. Personal style and social interaction skills may also have considerable stability after third grade, for Kagan and Moss (1962, p. 272) note that "the first four years of contact with the school and peer environments...crystallize behavioral tendencies that are maintained through young adulthood." Many workers including Kraus (1973) guess that a child's expectations are fairly stable by the end of third grade. One question the present research addresses then is: Do expectations become stabilized early in the school career?

Second, if one wishes to improve expectations, what paths of action are open? In the experiments where children's expectations are raised we show that positive evaluations by a suitable other are sufficient to raise expectations. This way of raising expectations was drawn directly from theoretical assertions and from closely related laboratory work with college students in expectation states theory. But we have no precise idea of how young children's expectations are raised or lowered naturally in kindergarten or first grade. If a child gets a low mark in conduct, does this have any noticeable effect on his expectations for reading? Or if a child has higher expectations for his own performance than his parents hold for him, do his expectations decline or do his parents' expectations increase? To change children's expectations by having some kind of "expectation raising" session fashioned after our experiments might at times be feasible for children with unrealistically low expectations. One would think it preferable, however, for a child to begin school with high expectations for himself and then to have those expectations confirmed rather than to try to elevate expectations after they have been pushed down. It may be next to impossible, for example, to improve a child's expectations for academic success by the time he has reached junior high school if he has experienced little success up to that time.

The reason is clear. Once the self-fulfilling prophecy is put in motion it tends to perpetuate itself. If a child's first report card rates him low on reading, his expectations may be shattered. In addition, hw may then be assumed by his teacher to "be slow," and given little time or opportunity to prove himself. His classmates may not pick him to be a participant in games or contests including reading. Our experimental data (Webster & Entwisle, 1974) show clearly that expectations influence evaluations and the same performance gets different evaluations depending upon the presumed ability of the performer. Also, Seaver's (1973) study shows that expectations first-grade teachers hold because a child follows his "bright" or "dull" sibling in the teacher's class are important determinants of evaluation. There is some limited evidence that expectations for the future reflect initial levels of success (Adelman, 1969) and also future achievement itself (see Finn, 1972). Because of one event, say a first mark that happened to be low, both the way a child is treated and the kind of treatment he comes to expect may be altered. Such events could have profound implications for a child's academic career.

Two kinds of effects emphasize the importance of expectation setting. First, considerable evidence exists (Rist, 1970; Palardy, 1969; Lambert, 1970) that teachers' expectations are not "open to evidence." Some teachers (see Palardy, 1969) expect boys to do as well as girls in reading and boys in their classes do tend to do as well as girls. How children behave or perform, in other words, may depend less upon their natural talent and effort than upon the expectations that shape teachers' responses to them. Second, although teachers can form expectations on other grounds than objective test scores, as the Rist (1970) and Williams' (1970) studies show, there is only a slight relation, if any, between IQ's obtained early in the school career and early performance levels. Forming expectations on the basis of IQ test scores may not be any fairer. As the reader will see, when first-grade middle class children in this study are sorted into quartiles according to IQ test scores, there is no relation between IQ and early reading performance. Thus it does not make sense to recommend that teachers group children on the basis of tested IQ rather than body odor--it might not be any "fairer." Rather each child is entitled to have his high expectations not destroyed so they will lead him to perform and to learn.

There is considerable confusion about subverting "academic excellence" and the use of ranked grading systems. The seventh-grader who performs well is likely to be the child who has performed well in sixth grade, fifth grade, and all the way back to kindergarten. But surely the curriculum of the elementary school is within the mental grasp of all children except those with marked and pathological retardation. There is no reason to think any child cannot learn the multiplication tables, or considerably more complex materials, under "proper" conditions. Proper conditions may include having expectations for a reasonable degree of success.

The Social Context

There are two major lines of research that point to the importance of the social context in measuring expectations. One is the perplexing and often contradictory results in surveys of the level of self-esteem of black or disadvantaged ckhildren, viewed in the perspective of Rosenberg and Simmons' large cross-sectional survey over several contexts. The other is a conflict in findings between some experiments of Cohen's and of our own concerning how black children perform in mixed-racial interaction.

Some studies of social interaction in mixed racial groups of children, all with older children, indicate that blacks suffer from interaction disability (see Cohen, 1972; Cohen, et al., 1972; Cohen and Roper, 1973; Katz and Benjamin, 1960; Katz, Goldstein, and Benjamin, 1958). These studies concern older subjects (junior high age or older) than those in our experiments and the findings run somewhat counter to what we observe in our experimental studies with younger children. We did not find that black children who attended an integrated (60% black) school

from kindergarten on were disabled in interaction with whites in mixed racial work groups (see Entwisle & Webster, 1974b). We found comparable volunteering rates and equal susceptibility to expectation raising. It is important to study how young children's self-expectations develop in natural settings for two reasons, (1) because our observations of blacks in mixed racial groups differ strikingly from Cohen's and (2) because no one has ever before looked to see how expectations develop starting at the time when the child enters school.

A word is in order here concerning the use of the concept expectations in this analysis. By "expectations" the authors mean a realistic guess of an evaluated performance. In particular a child's expectation refers to the child's guess about the mark he will receive--his scholastic performance as evaluated by the teacher. A parent's expectation likewise is the parent's guess concerning the teacher's evaluation of the child's scholastic performance, the mark his child will receive. The terms "forecast" and "hopes" occasionally appear in this analysis in place of "expectations" where they provide an easier intuitive grasp of the data; but by such terms the authors do not intend anything other than "expectations."

Self-esteem, which one might term the global set of self-expectations and the affect surrounding them, has so far been difficult to measure. A fundamental stumbling block is the lack of a handy validating criterion. Who but the person himself is party to his feelings of self-confidence, his pride in ownership or accomplishment, his hopes for the future, and his memories of past successes or failures? What indicants of these exist outside the person himself? If in a classroom a child raises his hand often to volunteer answers one might think the child had high self-esteem. Unfortunately, however, volunteering may indicate a need for attention or a need to impress classmates just as well as self-confidence. In the case of intelligence, defined narrowly for purposes of test construction in terms of performance in school, a handy measuring rod exists. But self-esteem has no such handy yardstick. It is hard to specify criterion behaviors or the evidence to use in measuring self-esteem. Dickstein (1972) sensibly chose a multiple criterion, peer popularity plus academic performance, and found validity coefficients in the range from .40 to .50. Such validity coefficients are much lower than those customarily found for cognitive tests. It would probably be a mistake, however, to insist that a self-esteem measure correlate more highly with this multiple criterion, for self-esteem obviously has many more facets than the two represented in the criterion.

Self-esteem, as William James pointed out, can be thought of as the quotient of the individual's successes compared to his "pretensions." Another way to express this is to define self-esteem as the discrepance or residual between achievements and expectations. If a person has high expectations and achieves well, then the discrepance is small. This kind of person will have higher self-esteem than the person with high expectations who achieves average or below and has a large discrepance. On the other hand if the person has modest expectations and achieves at a high level one would expect his self-esteem to be higher than the person who

achieves high against a background of high expectations. Such analyses, although they may have some relevance, need to be carefully balances in terms first of the individual's information-processing ability and second in terms of what is "high" and "low."

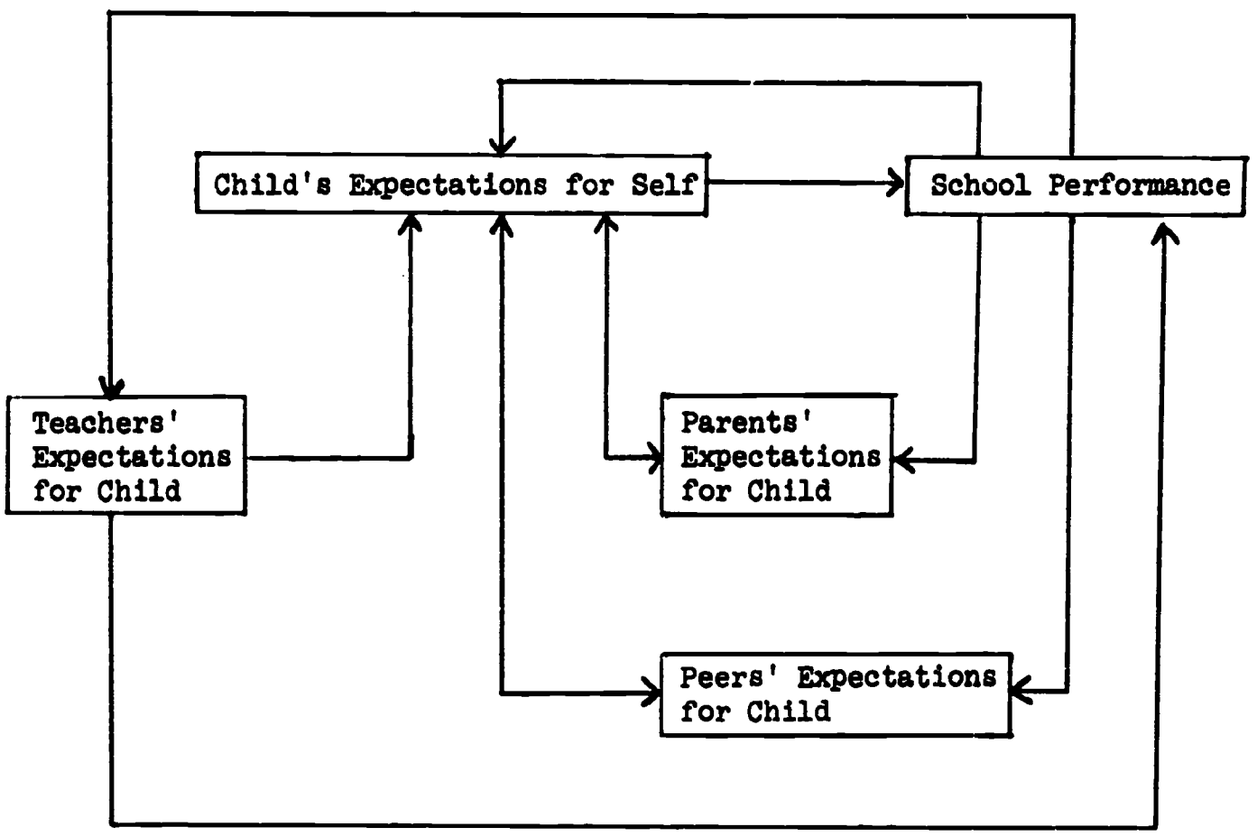
Persons process information about themselves to form an evaluation. Becuase of a possible need to preserve and enhance the self, negative information is downplayed or perhaps even ignored. Thus a child who does poorly in arithmetic may literally lose sight of evaluations in this area, or turn his attention to other "more important" areas. There is some evidence of this in the Rosenberg and Simmons study where black children with low IQ's actually report having parents who see IQ as less important for schoolwork than children with higher IQ's. Self-esteem as such may be undifferentiated in young children and based on factors that the group as a whole does not see as value. A young child as we know mainly from observation may place great value on an object that no one else values. He similarly may regard himself or his own characteristics in ways other persons do not share. He may be ill at ease because he has freckles, or because his speech differs in small phonological ways, or even because he is later in losing his deciduous teeth than his classmates. It takes some time for experience to accumulate if he is to discover what the relevant patterns are and how he stands on them. Thus self-esteem early in the school career may easily be "lop-sided" or grossly inflated or deflated. As time passes however the child's self-esteem may converge toward an evaluation which would coincide with an evaluation made by significant others. In the early stages however his expectations for himself in various important areas are probably not well-integrated. Certainly for school subjects and for peer-group performance he often has insufficient data to justify a stable estimate. Social context can be regarded as a feedback generator, but the amount of feedback obviously depends on how long the child has been in that context.

What is "high" and what is "low" depend entirely on context. A child in a lower-class school who receives an A may be performing at a high level compared to other children in his class. Still in absolute terms he may be performing poorly compared to children in a suburban school or compared to national grade norms. This is so obvious that it hardly needs to be said yet in studies of self-esteem in older children comparisons have been made across groups ignoring social context. To use a physical analogy, this can be likened to saying that an object that weighs one pound on earth weighs one pound on the moon.

The Model

Bronfenbrenner (1973) has recently called for research on children that embeds the child in a social context and that examines feedback. He also points to a need for research in naturalistic settings. The present research addresses both these matters. In addition, as the next chapter will specify at considerable length, the research is conceived in terms of a specific causal paradigm. The aim is to specify components

Figure 1.1



The Dynamic Model

of a model, eventually specifying the components in a precise quantitative manner. The data analyzed for this report, although voluminous, are still not voluminous enough to allow many precise quantitative statements. On the other hand, they do suggest that such statements could be made by adding more data of the same kind, a task we are presently pursuing.

Studies of status attainment in adolescents see the individual nested in a context of significant others where parental pressures, peer friendship commitments, and school (teacher) factors operate to shape vocational or educational aspirations. Our model of the expectation-shaping process in young children is reminiscent of those models, with the additional complication that we measure variables at several points in time.

The model underlying the design and analysis is a dynamic one. It is thought that the child's initial expectation level, shaped by home factors before he comes to kindergarten, and little influenced by "objective achievement levels," is heavily but diffusely influenced by what happens early in school (kindergarten). Major focused influence is exerted in first grade as achievements are evaluated by his teacher and his peers, particularly in the prime academic areas of reading and arithmetic. In addition his parents, who heretofore have shown concern for his performance in non-academic areas, now indicate that they have expectations for his performance in school, particularly in substantive areas like reading and arithmetic. As he performs and is evaluated, peers, parents, and teachers modify their expectations for him and convey this information to him by way of evaluations. The diagram below summarizes the feedback processes believed to shape expectations. The process portrayed by the model operates <u>over time</u>. The time frame for measuring variables implicated in the process will be indicated in the data collection schedule. The principal advantage in using a longitudinal approach is that it permits the teasing out of causal patterns. Only by observing that some consequences follow particular antecedents can one isolate causes. <u>Changes</u> in level, i.e., dynamic rather than static effects, are the main targets of the analysis. As the reader will see, discrepancies in expectations and performance are what lead to action or change, not the absolute level of expectations <u>per se</u>. Failure to observe self-image in this way is probably one reason previous work has led to confusing results. Also, as mentioned earlier, expectations or self-image can be measured only in a social context, an approach previously overlooked. Expectations are of necessity relative.

Studies by Brookover and his associates (1962, 1964, 1965, 1967) show that older children have different "self-images of ability" for different substantive areas, and our work here shows that first-graders have very different expectations for reading and for arithmetic. Accordingly, one can imagine the model operating for a child in reading, separately in arithmetic, and perhaps also separately in conduct, the three areas for which marks are collected from school records. The feedback process probably operates independently for the child's perceptions of his own sociometric standing as well.

Children are drawn from three residential locales whose characteristics might cause the variables in the model to have different impacts (i.e., the path coefficients may differ depending on the social setting). Black children in an all-black context, for example, may respond to peer expectations more (or less) than black children in an integrated setting. Or white children with a black teacher may respond less to teacher evaluations than white children with a white teacher. (Some limited data for black junior high school students show less influence of peers' upon black children's expectations and a weaker relation between ability self-concept and performance for blacks than for whites, although for both races self-conception of ability was a better predictor of performance than IQ. Morse in Brookover, et al., 1967.) Thus the dynamic model could operate fairly independently for the same child depending upon whether performance in reading, arithmetic, or conduct is involved. Also the model could assign different average values to factors like "peers' expectations" depending upon the broader social context of the child's social milieu. The model, in other words, may specify feedbacks to the child of different inportance depending on the child's social class or racial context.

It is important to notice that by aggregating over several first-grade (or second-grade) samples from the same context a sufficiently large set of data can be obtained to perform the planned analyses. The writers feel that this aggregation is superior to aggregating first-grade cohorts from different schools because of the homogeneity of the samples from the same school over short (2 or 3 year) time spans. So far there are two first-grade cohorts aggregated for the middle-class school which will be analyzed in Chapter 5. The lower-class school has only one cohort so far and results for it are given in Chapter 6.

CHAPTER 2
METHOD AND DESIGN

The Conceptual Framework for the Design and Analysis

Psychological and Sociological Paradigms

One way to conceptualize educational research is by the unit of analysis. Traditionally educational psychologists have analyzed individuals, for example, how a person's perceptual skills affect his reading achievement. Sociologists of education, on the other hand, have analyzed social aggregates, addressing such issues as the school performance of subcultural or minority groups. Both lines of inquiry look at outputs, achievements of the individual or of the social group. Both tend to skip over how inputs are converted into outputs of achievement, in particular what interpersonal events foster achievement. If middle-class children read better than lower-class children, exactly what happens day by day in a middle-class setting that brings about superior reading? What social processes or activities occur to account for the differences?

This research tries to trace out how sequences of events or patterns of social interaction lead children to take particular views of themselves as performers in school. The unit of analysis is the school child nested in a group of significant others, his parents, his peers, his teacher. This unit bridges both the psychological and sociological units of analysis mentioned above, and focuses on the articulation of the child with his group of significant others. The aim is to trace out how social factors and processes are translated into expectations of children and how these expectations, in turn, affect children's performance. The design and analysis draw upon both the psychological and sociological tradition of research.

Like longitudinal research in psychology, children are measured at several time points. The child's expectations for himself are assessed repeatedly from a time before he receives his first report card to the end of the second grade in some cases. Individual curves can be derived which show how a child's hopes for his own achievement change over time, indicating whether, for example, a given child's expectations remain stable or move. With expectations rising on the average over the first-grade year (as they do for instance in the white middle-class school) there is opportunity to see whether all children's expectations rise a little, or whether some rise a great deal while others decline slightly. It could turn out that a child's expectations decrease if his earliest expectations are not confirmed, or equally well that his expectations remain fixed and his performance improves.

The most obvious advantage of longitudinal data is the one hinted at just above: it does away with the confounding inherently present in cross-sectional data and in aggregated data. An average gain can imply

many different kinds of performances at an individual level. As has received much attention in the psychological literature, however, longitudinal data need careful study to separate out (1) effects of age, (2) effects of cohort and (3) effects of time of sampling. An example will make this clear. A child who is 5 in 1970 is obviously different from a child of 6 in 1970 because of attained age, but also because the 5-year-old is drawn from a 1965 cohort and the 6-year-old is a member of a 1964 cohort. In addition data from both children, being taken in 1970, may differ in important ways from data procured in 1971 from 5 and 6-year-olds at that time. The conditions of study in 1971 differ twice (sampling difference) and the events intervening between 1970-71 differ from those between 1969-70 (cohort differences). So effects of age, cohort, and time of sampling all need specification. Fortunately in the present research the time-span is relatively short and the same school and neighborhood are used for different cohorts. It is therefore assumed that differences between first and second graders of the same cohort are largely age or experience differences rather than differences stemming from time of sampling. By looking at cohorts separately and neglecting time-of-sampling differences--looking at, for example, first graders in the same school in 1971 and 1972--to see if samples of the same attained age differ, we can estimate the size of cohort differences. As the reader will see, cohort differences, insofar as they are available (for one school only) appear small. This fortunate state of affairs allows cohorts to be combined and, neglecting time of sampling on a priori grounds, we then attribute changes to differences in age or maturation.

The research also draws from the sociological tradition. Like sociological studies, this work studies contextual variables and tries to impose a causal paradigm. Children's expectations at the end of first grade, for example, are studied in the context of (1) expectations of the child earlier that year, (2) parents' expectations, and (3) previous marks received. The number of cases involved (so far 150 at the maximum) and the nature of the measurement (usually ranks) does not permit a full correlational or path-analysis type of treatment at this stage in the research. However, repeated cross-tabulations are carried out where change in one variable is tabulated against change in another variable. Such an analysis can examine whether children with low initial expectations revise their expectations upward after receiving high marks, or whether marks and expectations change simultaneously so as to increase the agreement between them from one time point to the next. Or, to take another example, if parents on the average expect boys to do more poorly than girls in reading, what is the course of particular boys' expectations whose parents hold very low expectations and how are the expectations of both parents and boys affected by marks the boy receives? Or, to take still another example, if black children have higher expectations at the start of school than white children, is this linked to higher expectations of black parents, black teachers, or peers for these children, and does this elevated state persist?

Sociologists have taken particular interest in status attainment by adolescents because of its obvious links to classical problems in sociology, how social stratification is effected and what makes it persist, for example, or the converse problem of how and why social mobility occurs.

This research tries to investigate variables similar to those in the status attainment work but our data are drawn earlier in the life cycle. The principal "attainments" are performance in the core areas of reading and arithmetic. The present research includes "significant others," parents, peers, and teachers, and the influence exercized by each is assessed. This procedure closely parallels the sociologists' strategy with older individuals. Also in parts of our analysis there in no intent, as in the psychological-type studies alluded to earlier, to follow individuals one by one or to trace out individual paths. Rather, like sociological studies of status attainment, the overall thrust is to evaluate unique variance explained and to assign direction of causality between variables. The main difference between the present study and the usual kind of sociological analysis is that attainment level is measured several times (twice each school year), so the model is dynamic.

The melding of the psychological and sociological approaches found in this research is novel and allows study of the dynamic causal model depicted in Chapter 1. The longitudinal method is often extolled because it allows one to observe change in behavior as the individual grows. Obviously, however, time cannot be viewed as the independent variable in the sense of causation. Rather this method allows study of sequential intra-individual variation. If hopes of children in one cohort assessed at the end of first grade are weighed against hopes of other children in second grade, one might conclude, other things being equal, that the average level of hope had risen or fallen. One might also calculate the correlation between hopes and marks within the first-grade sample or within the second-grade sample. One could not, however, conclude that high hopes in the second grade were preceded (or caused) by high marks in the first. Thus the advantage of the longitudinal method lies in teasing out causal patterns not in detecting successive differences with greater power, an aim often cited by developmental psychologists and by sociologists who carry out parallel studies.

Also it should be noted that aggregation of some kinds will be carried out--similar cohorts can be combined--but other kinds of aggregation will be avoided, largely on a priori grounds. Separate analyses, for example, will be carried out for samples from segregated and integrated schools. Also in many instances, sex effects will be searched for by analyzing data for boys and girls separately. This kind of preservation of context and attention to the cluster of significant others leads to the discussion in the next section of nesting of subjects.

Early Stages in the Life Cycle

The present research measures variables in young school children which may be potent determinants of later achievement and of ultimate status attainment. To our knowledge, these variables have not previously been studied in a sociological-type causal framework. Several studies suggest that by the end of third grade the prediction of eventual attainment may be almost certain. Both national (Bloom, 1964; Kraus, 1973) and cross-national studies (Husen, 1969) show very little change in children's achievement levels in school after that time. These findings, it should

be noted, are in accord with much folk wisdom as well. If this kind of early asymptote then exists, two conclusions may be drawn. (1) Status attainment may be predictable from points very early in the life cycle. Despite data showing that peers and the social context in high school may account for considerable variance in later attainment, these peers may be surrogates in a sense, carriers of influence for peers who intervened much earlier. If a person's level of educational attainment is set by the end of third grade then perhaps his peers before that time have shaped his tendencies to achieve. His peers later, in high school, are "influential" because they match his earlier peers in their characteristics. It is only because he remains in a fairly constant environment that both sets of peers appear equal. (2) Policy-oriented or action research should perhaps focus on school children at a much earlier point in the life cycle, before the point when eventual attainment becomes predictable. There may be little use in altering school or school environments after the early primary years if educational attainment levels are as stable as the studies quoted above suggest.

The generally equivocal and discouraging results from Headstart not withstanding, it may be that the primary school years are crucial. For one thing, the Headstart programs concentrated on cognitive variables, langu[illegible]ls and the like. There was almost no effort directed at affective a[illegible] motivational variables, and little attempt to assess such variables. Expectations are a kind of affective variable. Early in the life cycle before much feedback has occurred, they may be very influential, and independently influential. Later expectations may covary strongly with cognitive variables because the kind of performance the child displays will lead to various kinds of evaluations. A child by the end of third grade may be fairly good at forecasting his own performance; his expectations then may reflect less his positive or negative feelings toward himself and the subject matter than his recollections of what has happened in the past. Also teachers may not evaluate him on a "clean slate" at each evaluation opportunity but come to expect a certain level of performance and "see" that level of performance no matter what the child does. (A recent study reveals that teachers even carry over a "slate" from one sibling to the next sibling in the family (Seaver, 1973).) A number of considerations suggest that <u>very early</u> in the school career a child's experience could be crucial.

The Importance of Social Context

The social context germane for this longitudinal study derives from two separate theoretical orientations. The first, the nest of significant others, has already been linked in Chapter 1 to the developmental "looking-glass" of Mead and Cooley and to Sullivan's set of "significant others." The set of significant others has been an important and continuing concern in sociological studies of status attainment. The only difference in its use here is its application to students who are much younger than those previously studied and who have just emerged from the protective circle of the family.

The second theoretical strand, more macro-sociological in its roots, links this research to the stratification system of the society as a whole. The cross-sectional study of social class differences is firmly established as a research paradigm in both education and sociology, for example studies of school achievement over various levels of social class. The repeated demonstration of social class differences has reached a point of diminishing returns, however. Middle-class children are on the average better readers at every grade than lower-class children. This fact is well established. What is not known is the causal priority to be assigned to social factors leading to the association between social class and performance. One large and controversial question at the moment is the causal priority of genetic or environmental factors in school achievement. The authors wish to sidestep this problem and concentrate on other problems less sharply in focus at the moment, but perhaps equally important.

What are some provocative and counter-intuitive findings that could shed light on causal priorities of social factors? An outstanding recent example is the shower of studies showing that low SES or minority children do not have lower self-concepts than middle-class or majority-group children. This finding has been observed often enough so it is unlikely to be overthrown. It was unexpected because workers assumed that the self-view of minority-group children would reflect the view that the society at large held for their minority group. Thus adult blacks enjoy relatively low prestige in the larger society and so it was assumed that they--and black children--would hold low views of themselves as individuals. This picture now seems quite false, at least as far as children are concerned.

Another exception, less widely known, is the finding by one of the authors (Entwisle, 1968) that inner-city (disadvantaged) first-grade children of average IQ have developed linguistically to a point attained by suburban first graders in the gifted IQ range. This finding has not been independently replicated and so is not as firm as the finding of high self-esteem for minority-group children, but it holds up over more than one measure in the same study and is consistent between black and white sub-samples within the study. These two counter-intuitive findings--high self-esteem of blacks and advanced linguistic development of black first graders--underline both the importance of events early in life and the need for social context to be preserved.

Rosenberg and Simmons' (1971) work is especially informative with respect to the impact of social context. In a large cross-sectional study of children at third-grade and higher, they note that minority-group children in certain contexts have higher self-esteem. The overall picture for minority-group children is for higher self-esteem because most of them are in consonant social contexts which protect self-esteem. The black child in an all-black neighborhood and all-black school compares himself with those around him. He does not use children or families with which he has no contact as reference groups. For this reason, if his parents are separated or divorced or if his actual father is unknown, his family status is not a source of embarassment or sorrow because many of his confreres have the same problem. A black child from a broken family

who attends an integrated school, however, rates himself in comparison with his classmates whose families are intact. In this atmosphere a broken family produces social stigmata.

The impact of other facets of school life that might also be thought to lower self-esteem is cushioned by self-reference to protective groups. The low-IQ black child, for example, reports that his parents believe that being smart is not very important to getting higher grades. Or, for another example, the black child with very dark skin rates color as less important in assessing physical attractiveness than the child with light skin color.

The important outcome of Rosenberg and Simmons' work for the present research is that self-esteem, which can be termed a "global set of self-expectations," is highly dependent on context. Rosenberg and Simmons could not investigate causal priorities directly because their data are cross-sectional, but their data (retrospective reports of racial taunts) implicate the kinds of interpersonal events that could lead to lower self-esteem in minority-group children. Certainly racial taunts are more likely in a mixed than in a segregated context.

Social context and its impact on expectations has also been pointed up in small-scale experimental work. The study of interventions to raise expectations of black children in small mixed racial groups (see work by Cohen and her associates) when balanced against some small experimental studies of our own (Entwisle & Webster, 1974a, 1974b) point even more strongly to contextual effects and their importance. In brief, in mixed racial work-groups of children brought together for experiments, Cohen found that black children held low expectations for their own success and that these feelings were shared by white work-mates. Blacks who up to the time of the experiment were resident in all-black contexts did poorly in the mixed-race experimental context even when much preparation before the experiments guaranteed the superior competence of the blacks. Blacks were given intensive crash-course training on assembling radios or the other tasks to be carried out with whites in later experiments, while whites had no such preparation. Cohen's experiments document the low expectations held by white children for black children and that these low expectations impair blacks' performance.

Guided by the importance of social context in both the large-scale Baltimore study of Rosenberg and Simmons and the small-scale experimental work, the present research focuses on children within two contexts--one all-white and one mixed racially, 60% black. (Later work will examine all-black contexts.) This research can examine more events, and more specific events, than that of Rosenberg and Simmons and also, of course, looks at effects over time. If a child is asked to recall racial taunts, as in the Rosenberg and Simmons study, such a report has weaknesses too well-known to be enumerated here. Without being at all critical of the Rosenberg and Simmons work, one can note that self-esteem covers many aspects of the self. To understand its ups and downs or the underlying factors which comprise it, considerable precision is required. The causes of self-esteem are no doubt manifold and subtle, and a single overall

measure is probably inadequate for explaining its causes. The present research attempts to look at the course of development of the academic self-image in particular areas, how performance feedback shapes that self-image, and how feedback from particular persons operates on the self-image. The social context in the persons of particular actors shapes the image, subject to overall constraints in the social milieu. Specifically, peer feedback in an integrated environment may have effects different from such feedback in a segregated environment. From a methodological standpoint it would be fruitless to try to guage the self-image with either the localized or general context erased. Up to this time studies have not combined both social context and a detailed dynamic feedback mechanism.

Description of Procedures

Three cohorts of children have been followed from the time they began first grade. Two cohorts are in a white middle-class suburban school. One (S-1) began first grade in 1971, the other (S-2) began first grade in 1972. A third cohort (L-1) in a mixed-race (60% black) urban lower-class school began first grade in 1972. Insofar as possible, similar data and information were obtained for all three cohorts.

The time chart below shows the sequence and timing of several of the repeated measures over a 2-year time span. Only cohort S-1 has been followed for the entire two-year period. Cohorts S-2 and L-1 have been followed on the same plan as S-1 but for only one year. The measures identified on the time chart will now be described.

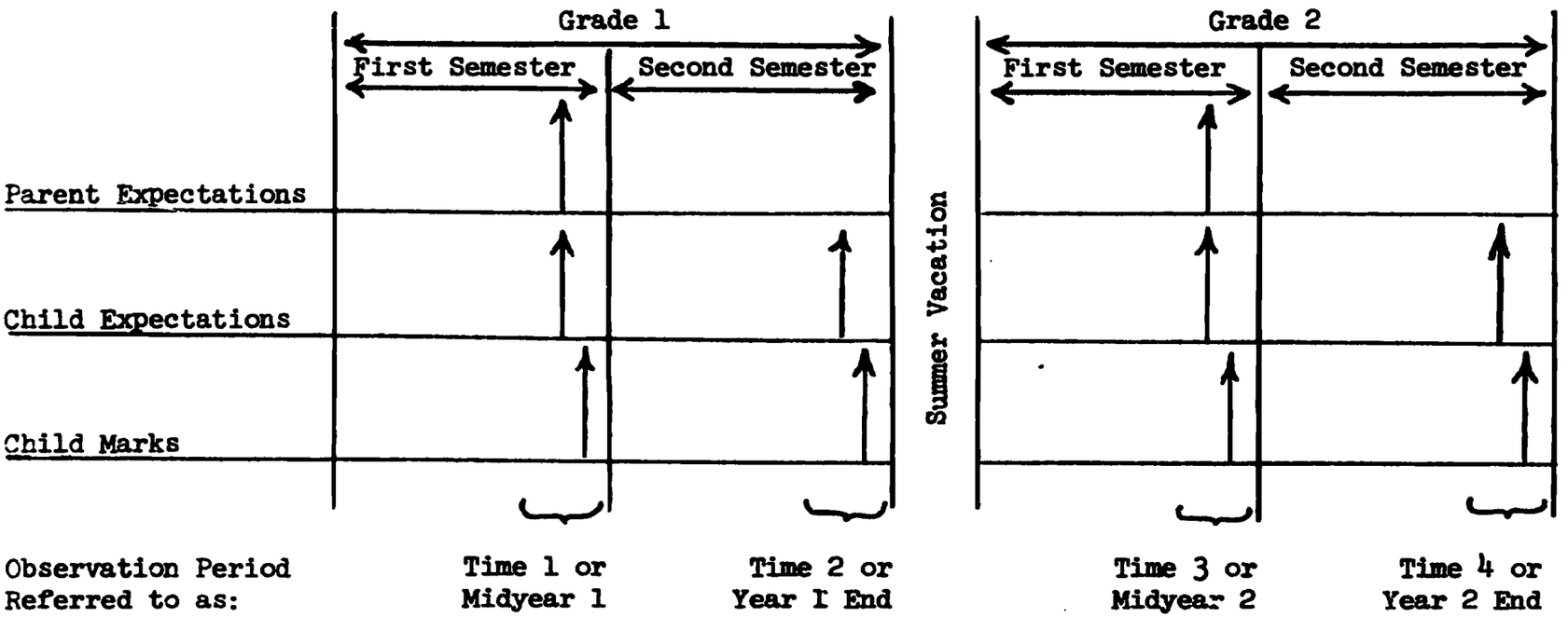
The Children's Expectation Measure

To provide a measure of the child's expectations for his own school performance, children were asked to "guess what your next report card will look like. Guess what you will get in reading...in arithmetic...and in conduct." How elaborate the interviewing procedure was which accompanied this "guessing" depended on whether the child was new to the study or had been interviewed previously.

For the initial measuring of expectations, a large plastic brightly-colored sheet (approximately 2' X 3') was prepared (see Fig. 2.2) with titles of school subjects (Arithmetic, Reading, Conduct) and squares for entering marks in it like a report card. This sheet was spread out on a table or sometimes on the floor. Next to the sheet were a number of piles of cardboard squares with large numerals (1, 2, 3, 4) inked on them.

Children, interviewed individuallly outside their classroom or in a separate room nearby, were told that "we are going to play a game--guessing what you will get on your report card." Before "playing the game," the child was asked if he knew what a report card was, what the numerals meant, and what "reading", "arithmetic", and "conduct" meant.

Figure 2.1
Observation Periods

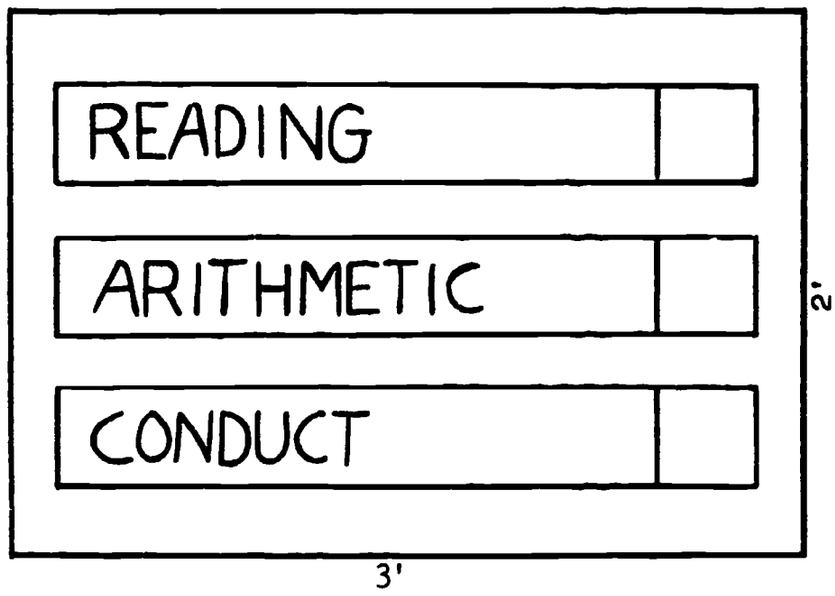


Notes:

1) an "↑" indicates observations were made on this variable (set) at the indicated time.

2) "Time 1" (stc.) refer to the Grade 1 - Semester 1 (etc.) observations made on a cohort no matter when the cohort was actually observed in terms of years.

Figure 2.2



Report Card Replica

Enough discussion then ensued so the interviewer felt reasonably confident the child understood what school report cards signified and how marks were coded. The child was then asked to pick a number from the pile of numerals and put it in a square next to "Reading" to "guess what you will get in reading." He was similarly asked to pick numerals to represent his guesses for arithmetic and conduct.* Initial interviews for measuring expectations in reading, arithmetic and conduct were held slightly before the child received his first report card in first grade. Report cards are issued 3 times in grade one, midyear, year end and half way between midyear and year end. Only midyear and year end report card information is used. In second grade report cards are issued four times--roughly corresponding to the end of each of four quarters of the school year.

As the child made his guesses the interviewer unobtrusively recorded check marks on a small 3" X 5" card. The cards were kept out of sight and it is doubtful if any of the children were aware their guesses were being recorded.

This same guessing procedure was carried out twice yearly during each school year. In first grade it occurred once before the child received his first report card (November for L-1, December-January for S-1 and S-2), and again before the end of the school year in lare May or early June. In second grade it was carried out just before the end of the first semester and again near the end of the school year.

As will become clearer from the analysis of the data, the expectation measure appears to have a fair degree of validity and reliability indicated by re-interviews, meaningful relationships that emerge between it and other variables, and a substantial degree of test-retest agreement when guesses for the same child are matched between one session and the next. A special reliability check run on a small sample of first graders with a short time span (one week between test and retest) indicates a high degree of concordance between expectations elicited on the two occasions (.76).

Sociometric Data

Two methods were used to procure children's sociometric standings within their own classroom. The first used a multiple-choosing procedure which forced a child to pick 6 children on two occasions (12 choices altogether) in such a way that choices would be widely distributed. The procedure went as follows. The class was randomly divided into three "teams," designated white, red, or blue. Then every child picked two persons from each team. The class was then redivided randomly into thirds. (One third of the former white team remained white but the other thirds were, respectively, now designated red or blue, with similar re-partitionings

---

*At a later time a separate interview was conducted with each child by a different interviewer. These interviews verified the child's understanding of marks and marking systems without reference to the guessing task. The independent verification is discussed in connection with results.

for the red and blue teams of the first round.) Again everyone picked two persons from each team. The results were tallied as the number of times each person was chosen and were later connected to the proportion of choices each child received. The scores were then normalized to be between zero and one. The normalization re-weighted choices to fall on a similar scale despite differing numbers of children per classroom. The multiple-choosing procedure just described was used only with S-1 in first grade.

The multiple-choosing procedure was too complex to be used with lower-class first graders. Another sociometric procedure was therefore developed and used with all classrooms starting in 1972 (S-1 in second grade, S-2 in first grade, and L-1 in first grade). The new procedure involved having a boy and a girl in each class designated as captains. Captains then took turns choosing their classmates to be members of their team in order to "play a reading game." Children to be captains were nominated by teachers. Teachers were asked to choose the boy and the girl in their class who were most proficient in reading and socially mature.

The experimenter, on entering the classrooms, explained to the class that "we are going to play a game" and called the captains to the front of the class without explaining how the captains had been chosen. As the captains took turns choosing children to be on their teams, a research assistant noted the order in which children were chosen. Children wore large numbers (like numbers worn by football players) so they could be easily identified. This procedure resulted in two separate rankings, one made for each captain's team, with the captain of each team being assigned the highest rank. In general the teams chosen were homogeneous by sex, girl captains choosing girls and boy captains choosing boys. The ranks by class were converted into scores between zero and one, so as to be comparable to sociometric scores obtained by the multiple-choosing procedure and also, of course, to normalize with respect to class size.

Sociometric rating scores between zero and one were assigned in such a way that the rating scores were equally spaced along the continuum between zero and one according to the rank order as obtained from the social choice ranking (i.e. order of being chosen or order in terms of proportion of choices received). Children tied on original rank order remained tied in terms of the resulting rating scale. Ties regularly occur in the "group-captain" procedure due to the parallel choice structure involved. The sociometric scale that results from the above procedure is a measure of the children's <u>within room</u> sociometric standing.

The sociometric ranking measurement was carried out once per school year per cohort at some time in the second semester. When the ranking was repeated with one lower-class group (20 children) on two occasions one week apart, the correlation between rankings was .73. The way the choosing occurs and the high correspondence one observes for the first half dozen children chosen for each team both suggest that there is a fairly clear, stable ordering for the top half of the class. For the bottom half there may not be much discrimination among individuals, but rather a generalized perception that the group as a whole is less able. The analogue of serial order effects appears to occur for children high on the list but not low. There does not seem to be a consensus as to who is "worst", "next worst", etc. although there is consensus as to who is "best", "next best" and so on down to about tenth in the class.

Data from School Records: Sex, Race, IQ, Absences, Standardized Achievement Test Scores, Teachers' Marks.

Sex and Race. These designators are self-explanatory except that all minority-groups except blacks (orientals and others) were classified as white.

IQ. In both schools, IQ tests are routinely given by school personnel with scores recorded in pupils' folders. For first graders in both schools IQ's are obtained using the Primary Mental Ability test early in the first-grade year.

Absences. Absences were tallied for the year in both schools on the report card issued at the end of first grade.

Standardized Achievement Scores. These will be available as students progress in school. The middle-class school does not have standardized achievement tests until the child is in third grade. We have none in data so far but will pick these up as they become available. In the lower-class school the practice of giving achievement tests has changed starting with the cohort under study. No achievement tests will be given until third grade.

Teachers' Marks. In both schools teachers assign marks in a wide variety of subjects (See Figures 2.3 and 2.4 for replicas of report cards in use in the middle-class and lower-class schools, respectively.). The reader should note carefully that the teacher's basis for assigning marks, according to the report card definition, is quite different in the two schools. In the middle-class school the teacher attempts to mark the child in relation to his own ability. In the lower-class school the teacher attempts to mark the child in relation to others of his grade level. Since both bases of assessment leave considerable room for interpretation, we queried teachers to get their definition of marking practices. Some written responses of teachers in the two schools, reproduced below, suggest that teachers do indeed attempt to operationalize the marking standards the report cards define. Samples of teachers' responses in each school are given verbatum below.
Lower-class school:

> Arithmetic. There are a list of math skills that each first grade must be taught. These skills are taught in sequential order... Diagnostic tests are usually given to evaluate how well the child has mastered a particular area....Those who mastered the skill received "good" (3). Those who even with additional reinforcement only got half on the test got 2 (fair) on their cards. Children who never mastered the skill got 1 (poor).

N.B. There seems to be confusion on this teacher's part about "1" being the highest mark--it was so defined on the questionnaire sheet containing the answer above and "1" also is defined as high on the face of the report card. Middle-class school:

R13955

BALTIMORE COUNTY PUBLIC SCHOOLS
PROGRESS REPORT

STUDENT _____ GRADE _____ SCHOOL YEAR 19____ 19____

TEACHER _____ SCHOOL _____

SUBJECTS	TERM 1		TERM 2		TERM 3		TERM 4	
	ACHIEVE	EFFORT	ACHIEVE	EFFORT	ACHIEVE	EFFORT	ACHIEVE	EFFORT
READING								
WORD ATTACK SKILLS								
COMPREHENSION SKILLS								
ARITHMETIC								
KNOWLEDGE OF NO. SYSTEM								
COMPUTATION								
PROBLEM SOLVING								
SPELLING								
BASIC WORD LIST								
OTHER WRITTEN WORK								
LANGUAGE								
WRITTEN EXPRESSION								
ORAL EXPRESSION								
SOCIAL STUDIES								
REFERENCE SKILLS								
BASIC UNDERSTANDING								

	TERM 1	TERM 2	TERM 3	TERM 4
HANDWRITING				
ART				
MUSIC				
PHYSICAL EDUCATION				
STUDY HABITS				
CONDUCT				

CODE: G - GOOD S - SATISFACTORY P - POOR

	TERM 1	TERM 2	TERM 3	TERM 4
TEACHER REQUESTS CONFERENCE				
IS BEING CONSIDERED FOR RETENTION				

JUNE 19 _____

PROMOTED TO GRADE _____

WILL WORK WITH GRADE _____

RETAINED IN GRADE _____

ASSIGNMENT SEPT. 19

ROOM _____

ACHIEVEMENT CODE (BASED ON GRADE STANDARDS)

A - SUPERIOR
B - ABOVE AVERAGE
C - AVERAGE
D - LOW AVERAGE (LOWEST PASSING GRADE)
E - DID NOT ATTAIN MINIMUM GRADE STANDARDS

"N" - INDICATES IMPROVEMENT NEEDED IN A PARTICULAR SKILL

EFFORT CODE

1 - MAXIMUM
2 - HIGH AVERAGE
3 - AVERAGE
4 - LOW AVERAGE
5 - MINIMUM

BEBCO 800 70

THIS IS YOUR RECORD. IT NEED NOT BE RETURNED TO SCHOOL

Figure 2.3

Middle-Class School Report Card

Figure 2.3 (Continued)

To the Parents:

This is the schools report to you on the progress of your child.

Reporting periods close on November 15, January 31, April 15, and at the end of the school year. Reports will be sent to you within a week of the close of each term. Your copy of the report need not be returned.

In the academic area your child will receive two marks - a letter grade (A, B, C, D, or E) for Achievement in relation to grade standards and a number (1, 2, 3, 4 or 5) to indicate Effort. "N" where shown indicates improvement needed in that particular skill. "D" is the lowest passing grade.

Note: Grade one pupils will receive no mark in spelling during the year.

The other areas, Handwriting, Art, Music, Physical Education, Study Habits and Conduct, are marked with a three point scale using G, S, and P as described.

A child being considered for retention at the end of the second or third term will receive a check (✓) in the space provided.

A space has also been provided for the teacher to request a conference. Please contact the teacher and arrange an appointment if indicated. You are invited, however, to ask for a conference at any time.

Joshua R. Wheeler
Superintendent of Schools

349 40 0151

BALTIMORE CITY PUBLIC SCHOOLS

REPORT OF _____ (LAST NAME FIRST)

SCHOOL NAME AND NO. _____

TEACHER _____

GRADE OR LEVEL _____ SCHOOL YEAR ENDING JUNE 19 _____

EARLY LEVEL

ACHIEVEMENT CODE

A – OUTSTANDING
B – GOOD
C – SATISFACTORY
D – POOR
U – UNSATISFACTORY

EFFORT AND HABITS CODE

1 – GOOD
2 – SATISFACTORY
3 – UNSATISFACTORY

	NOVEMBER REPORT	MARCH REPORT	JUNE REPORT
READING			
BOOK LEVEL			
ENGLISH LANGUAGE ARTS			
HANDWRITING			
LISTENING			
SPEAKING			
SPELLING			
WRITTEN ENGLISH			
MATHEMATICS			
HEALTH AND SAFETY			
SCIENCE			
SOCIAL STUDIES			
ART			
MUSIC			
PHYSICAL EDUCATION			

	NOVEMBER REPORT	MARCH REPORT	JUNE REPORT
CONDUCT			
EFFORT			
HEALTH HABITS			
HOMEWORK			
SAFETY HABITS			
WORK HABITS			
TOTAL TIMES LATE TO DATE			
TOTAL DAYS ABSENT TO DATE			
TEACHER REQUESTS A CONFERENCE — DATE			
TEACHER REQUESTS A CONFERENCE — TIME			

	YES	NO	YES	NO	YES	NO
TEACHER COMMENTS ATTACHED						

PUPIL MAY NEED MORE TIME AT THIS LEVEL _____ [ ] [ ]

ASSIGNMENT NEXT YEAR _____

TEAR ALONG THE DOTTED LINE AND RETURN THE LOWER PORTION

PUPIL'S NAME _____ ROOM NO. _____

_____ I HAVE STUDIED THIS REPORT.

PARENT'S SIGNATURE _____

SCHOOL NAME AND NUMBER _____

CONFERENCE SECTION

_____ I SHALL BE ABLE TO KEEP THIS APPOINTMENT.

_____ I SHALL NOT BE ABLE TO KEEP THIS APPOINTMENT.

I SHOULD LIKE TO HAVE A CONFERENCE WITH YOU ON

_____ DATE OR _____ DATE

Figure 2.4 Lower-Class School Report Card

183/184

A Few Helps in Understanding This Report

Your child's progress is reported in terms of his achievement in certain subject and habit areas. He is being measured in terms of his progress in reaching standards or levels that are considered appropriate for his age or years in school.

The letter "D" is used to show a barely satisfactory level of work. No parent or child should be satisfied with this grade. The child should strive to do work that is satisfactory or better.

Reading is a complex skill that involves the use of many types of materials. Some books show a grade level. Other materials do not.

The book level is indicated to let you know what book your child is handling in the classroom. These are the symbols used for readers:

PR – Material a child is using in getting ready to read

PP – Pre-primer material or first printed material in book form that is introduced to the child.

P – Primer material or material that gives many reading experiences before use of the first reader

1 – First reader } Publisher's grading of
2, 3, 4, 5, 6, 7, 8, etc. } a book

For further information about your child's progress in reading or in any other area, you are urged to talk with his teacher.

You may also send your comments to the principal of the school your child is attending or to the Elementary Division, 3 E. 25th Street, Baltimore, Maryland 21218.

Baltimore City Public Schools
THREE EAST TWENTY-FIFTH STREET
BALTIMORE, MARYLAND 21218

INFORMATION FOR PARENTS

Your child's success is not only important to you and your child but also to all of us who work in the Baltimore City Public Schools.

Three times during the school year, your child will bring home a report card. All you need to do for us is to cut off the lower portion, check that you have seen it, sign your name, and have your child return it to his teacher.

Look at the line Teacher Comments Attached. If the teacher has written yes, a special form will be stapled to the report card. If the teacher has written no, there will be no form attached to the report card.

You are encouraged to come in for a conference with your child's teacher. Each report form has a section where you may request a conference. Please use this section.

You are always welcome to come to school to talk about your child's progress.

Figure 2.4 (Continued)

When I assign grades for any subject for the first marking term of the year I... ook at the child in terms of his own ability. Our nongraded system enables us to mark a child in light of his achievements in regards of his capability instead of the traditional comparison to others who may have different capabilities....

If outstanding progress or 1 is to be considered then work must be:

a. consistent with ability
b. independent, self-directed, enthusiastic

If satisfactory progress or 2 is to be considered then:

a. child's work should be consistently accurate in subject
b. child lacks enthusiasm
c. child lacks self-direction in independent activities
d. child is still progressing

If moderate progress or 3 is to be considered then:

a. work is inconsistently accurate in subject matter
b. child shows an erratic pattern of effort
c. child lacks enthusiasm
d. child shows little self-direction
e. child is still showing progression

If not satisifactory progress or 4 is to be considered then:

a. child's work is inconsistent
b. child shows no independence
c. child shows no self-direction
d. child shows no enthusiasm
e. no progress is shown

I determine conduct marks by considering points:

1. consideration
2. interaction among peers
3. following adult requests

Self-Esteem

Self-esteem was measured on a scale especially developed for use with young children (Dickstein, 1972). Separate scales exist for boys and girls. Each scale contains three factors. These factor scores have been kept separate for the analysis, thus yielding three different indicators of self-esteem for each sex. The items (abbreviated) are shown for each factor, separately by sex, below:

Boys

I	II	III
polite	sports	arithmetic
obeying rules	strong	good student
cooperative	playing ball	learns new things quickly
kind	many friends	
helpful	gymnastics	
honest	running	
good student	right weight	
	able to look after others	

Girls

I	II	III
can look after others	sports	good-looking
can take care of herself	strong	many friends
polite	gymnastics	right weight
good student	playing ball	good student
learns new things quickly	running	polite
writing	dancing	
arithmetic		
honest		

Parents' Expectations and Questionnaire

Parents in the middle-class school were interviewed in many instances, the occasion being their appearance to visit their child's classroom during American Education Week (late October). Those not available were sent questionnaires by mail.

Parents in the lower-class school were also interviewed at school if they came for activities in connection with American Education Week. A much lower percentage attended however and the difficulties in securing the needed information from interviews in school led us not to mail questionnaires to these parents, but rather to send out interviewers to the homes to secure the information. Children carried home a note a day or two in advance of the interviewer's call notifying the parent that an interviewer would be coming, and explaining that this was a routine research request, not an indication that their child was in difficulty at school. Black adult female interviewers were despatched to homes of black children, and white adult female interviewers saw parents of white children. Refusal rates and other data are given with results (Chapters 3-6).

The data procured from parents was, insofar as possible, identical for the two schools. The parents were asked to "guess" what their child "would get on his next report card" and the interviewers recorded guesses on sheets of paper with replicas of part of the report card appearing upon them.

Analysis

All data were coded and punched on cards. Data reduction was accomplished mainly through use of standard DATATEXT programs. Further analyses and tests were calculated as needed. Selected results and the accompanying analyses are presented in Chapters 3, 4, 5, and 6, which follow.

CHAPTER 3*
MIDDLE-CLASS SCHOOL, COHORT 1
(STARTING FIRST GRADE IN SEPT. 1971)

Just north of the city line in Baltimore County, Maryland is a well-established middle-class neighborhood served by an elementary school. There are kindergartens in the school, two classes in the morning and two in the afternoon. There are three or four first grade classrooms, the number depending on yearly enrollment.

Tabular Summary

The first-graders who started school in September, 1971, at this middle-class school have been followed so far through two school years. The same kind of information was procured in both years. Data for the two years are summarized in Table 3.1.

Table 3.1 shows that in every area (reading, arithmetic and conduct) children expect on the average a B+ (numerical mean values range from 1.63 to 1.68) at midyear, and their expectations get higher at year end, especially in reading and conduct. Parents have lower expectations, averaging just slightly above a B (1.83 to 1.96). Parents' expectations in arithmetic are lower (2.03) the second year but reading and conduct stay much the same (1.87 and 1.90). Children's expectations consistently showed a decline between year end and the middle of the second year but a slight increase from the middle to the end of the second grade. By the end of the second year children's expectations averaged almost exactly what they did at the time they were first observed. (See also Table 5.1, columns labelled "Cohort 1" for a summary of means and standard deviations.)

The difference between the <u>average</u> expectations of parents and children does not give a very accurate notion of the difference between them, for the children's distributions are markedly skewed. For example, in reading 60 per cent of the children expect to receive the highest grade, whereas only 16 per cent of parents are that optimistic. At the other extreme, 6 children (7.5%) expect to get the poorest grade whereas only one parent (1.2%) forecasts a failure.

The variability of children's expectations is noticeably larger than parents' for every area at the middle of first grade but by the second grade, the variability of parents' expectations increased from first to second

---

*The reader may wish to read Chapter 5 first which presents combined results for Cohorts 1 and 2 in the middle-class school, and then return to Chapters 3 and 4 to pick up points not covered in Chapter 5.

Table 3.1

Means, Standard Deviations for Cohort 1
Middle Class School, 1971-72 and 1972-73

	Mean IQ = 115.44 S.D. = 10.13			Mean IQ = 104.1 S.D. = 11.47		
	1971-72			1972-73		
	N	Mean	S.D.	N	Mean	S.D.
Parents' Expectations--Midyear		(Time 1)			(Time 3)	
Reading	84	1.96	0.57	76	1.87	0.72
Arithmetic	84	1.83	0.60	75	2.03	0.77
Conduct	79	1.89	0.55	76	1.90	0.69
Child's Expectations--Midyear		(Time 1)			(Time 3)	
Reading	90	1.63	0.91	102	1.69	0.70
Arithmetic	90	1.66	0.90	102	1.77	0.90
Conduct	90	1.68	0.90	102	1.80	0.83
Child's Expectations--Year End		(Time 2)			(Time 4)	
Reading	90	1.42	0.60	102	1.63	0.63
Arithmetic	90	1.64	0.81	102	1.72	0.72
Conduct	90	1.49	0.72	102	1.68	0.63
Child's Marks--Midyear		(Time 1)			(Time 3)	
Reading	85	1.77	0.65	102	1.91	0.80
Arithmetic	85	1.82	0.49	103	2.07	0.62
Conduct	85	1.75	0.75	103	1.81	0.72
Child's Marks--Year End		(Time 2)			(Time 4)	
Reading	86	1.69	0.76	103	1.60	0.65
Arithmetic	86	1.65	0.59	103	1.94	0.64
Conduct	86	1.88	0.83	103	1.73	0.73
Mark Discrepance (Midyear Mark minus Year End Mark)		(T1 - T2)				
Reading	78	0.10	0.70			
Arithmetic	78	0.19	0.56			
Conduct	78	-0.04	0.63			
Expectation Discrepance (Midyear Expectation minus Year End Expectation)		(T1 - T2)				
Reading	85	0.24	0.88			
Arithmetic	85	0.00	1.14			
Conduct	85	0.21	1.03			
Mark-Expectation Discrepance--Midyear (Mark minus Expectation)		(Time 1)			(Time 3)	
Reading	82	0.18	1.07	96	1.91	0.80
Arithmetic	82	0.18	0.97	97	1.84	0.73
Conduct	82	0.10	1.06	97	1.79	0.84
Mark-Expectation Discrepance--Year End (Mark minus Expectation)		(Time 2)				
Reading	84	0.27	0.87			
Arithmetic	84	0.02	0.88			
Conduct	84	0.37	0.92			

Table 3.1 (continued)

I.Q. Correlations

PMA = Primary Mental Ability
SFTAA = Short Form Test of Academic Aptitude (from the California Test of Mental Maturity)

	1971-72		1971-72	
	N	r	N	r
Midyear I.Q.(PMA)-Mark Correlations				
	(Time 1)		(Time 3)	
Reading	82	.224*	79	.279*
Arithmetic	82	.022	80	.374**
Conduct	82	.109	80	.014
Year End I.Q.(PMA)-Mark Correlations				
	(Time 2)		(Time 4)	
Reading	82	.099	80	.145
Arithmetic	82	.008	80	.196*
Conduct	82	.043	80	-.001
Year End IQ(SFTAA)--Mark Correlations				
Reading			102	.131
Arithmetic			102	.201*
Conduct			102	-.010
Correlation of T1 I.Q.(PMA) with T2 I.Q. (SFTAA)			80	.655

* = p < .05

** = p p .01

grade while the variability of children's expectations declined from the middle of first grade to the end of second grade.

The teacher's rating of the child, i.e., marks, on the average, falls in between what parents expect and what children expect. In reading the average mark is 1.77 compared to 1.63 expected by children and 1.96 expected by parents. In arithmetic parents expectations and marks almost coincide (1.83 vs. 1.82). In conduct the spread is like that for reading.

The range of marks given by teachers is roughly comparable to parents' expected range but smaller than the child's expected range. The teacher gave no 4's and gave 85% to 90% A's and B's overall. Their marking standards in general then, are relatively easy. However, as will be pointed out below, the relative number of A's and B's is very different from one area to the next. About equal numbers of A's and B's are given in conduct and about 3 times as many B's as A's are given in arithmetic. Reading falls between, with about a 3:2 split favoring B's. Teachers then, vary noticeably in their marking practices from one subject to the next, despite the similarity in the average marks assigned.

The assignment of marks must be interpreted in light of the fact (see Chapter 2) that in this school children are <u>not</u> marked with respect to one another, but are supposed to be marked in terms of their own ability. If a child is judged to have "high" ability but does not, in the opinion of the teacher, perform up to his potential, then his grade is low. Theoretically, to take the other side of the coin, if he performs much better than the teacher expects, she should assign a high grade. From inspection of the actual mark distributions, children on the average appear to be slightly better in all three areas than the teachers expect.

Discrepances Between Expectation and First Report Card

Reading

In general, for children the first mark in reading must be surprising--in well over one-third of the cases (34 out of 82) an unpleasant surprise. In about one-quarter of the cases (20 out of 82) the child does better than he expects. All children who expect the worst (6) do better than they hoped. Perhaps the most comprehensive comment is that 66 expect <u>and</u> get A's or B's, but of these, 39 (59%) represent reversals--those expecting B's get A's or vice versa. There is <u>no</u> significant matching between children's expectations and marks at this early stage. Thirty matches are expected by chance, 28 actually occur. There is some indication that expectations in reading are higher than marks rather than lower in the cases of disagreement, but the trend is not strong enough to attain significance.

Arithmetic

In arithmetic there is likewise <u>no</u> significant agreement between children's expectations and the marks they receive. Matches would be

Table 3.2

		Reading Mark, Time 1				
		1	2	3	Total	Percent
Reading Expectation, Time 1	1	15	27	7	49	59.8
	2	12	12		24	29.3
	3		2	1	3	3.7
	4	1	4	1	6	7.3
	Total	28	45	9	82	
	Percent	34.1	54.9	11.0		100.0

Table 3.3

		Arithmetic Mark, Time 1				
		1	2	3	Total	Percent
Arithmetic Expectation, Time 1	1	13	35	2	50	61.0
	2	2	13		15	18.3
	3	3	10	1	14	17.1
	4		3		3	3.7
	Total	18	61	3	82	
	Percent	22.0	74.4	3.7		100.0

Table 3.4

Conduct Mark, Time 1

		1	2	3	4	Total	Percent
Conduct Expectation, Time 1	1	23	14	7	1	45	54.9
	2	10	13	3		26	31.7
	3	2	3	1		6	7.3
	4		4	1		5	6.1
	Total	35	34	12	1	82	
	Percent	42.7	41.5	14.6	1.2		100.0

expected 28% of the time by chance, 33% actually occur. Children were optimistic about arithmetic to start with for 61% expected an A. The outcome must have been disappointing, however, because the teachers are noticeably less generous with A's (22% of grades are A's in arithmetic). Thirty-seven children (44%)--or almost half the total--do not do as well as they expected, and the tendency to do worse rather than better than expected is significant ($\chi_1^2 = 6.56$, $p < .05$).*

Conduct

Teachers apparently are most lenient in their grading of conduct--giving about 85% A's and B's and dividing marks about equally between the two. It is about twice as hard to get an A in arithmetic as in conduct, with reading standing in between. Again the amount of matching between children's expectations and their marks is not significant. Although there is some tendency for expectations to exceed marks when disagreement occurs, the tendency for over- compared to under-estimation is not statistically significant.

Discrepance Between Expectations and the Report Card at the End of First Grade

Reading

The year end report card must also have been surprising to children (Table 3.5). Less than half (45%) got the mark in reading they expected and this proportion is within the range of chance. Of those 55% who received a mark different from what they expected 72% did not do as well as they hoped. Of the remainder almost all (24%) were children who expected a B but got an A. The tendency for children to expect a higher mark than they received is significant ($\chi_1^2 = 8.69$, $p < .01$).

If anything, children's expectations for A's and B's are <u>more</u> polarized on the year end report card than on the first--98% now expect a B or an A compared to 89% earlier. This means that even though more A's are actually awarded by the teacher (48% vs. 34% awarded earlier) children's expectations have increased enough in the interim so that about the same degree of negative feedback ensues.

---

*χ_1^2 values calculated for the comparison of cases above and below the main diagonal have not, in general, been corrected for continuity within this report. Since the expected values are quite large (20 or more) such a correction would have a negligible effect. (See G. Snedecor and W. Cochran, <u>Statistical Methods</u>, Sixth Edition, pp. 209-213 for a discussion of the relationship of the χ_1^2 correction of this type to the usual continuity correction for the normal approximation to the binomial.).

Table 3.5

		Reading Expectation, Time 2					
		1	2	3	4	Total	Percent
Reading Mark, Time 2	1	28	11		1	40	47.6
	2	20	10	1		31	36.9
	3	4	8			12	14.3
	4		1			1	1.2
	Total	52	30	1	1	84	
	Percent	61.9	35.7	1.2	1.2		100.0

Table 3.6

		Arithmetic Expectation, Time 2					
		1	2	3	4	Total	Percent
Arithmetic Mark, Time 2	1	20	13	1		34	40.5
	2	20	18	5	2	45	53.6
	3	3	1		1	5	6.0
	Total	43	32	6	3	84	
	Percent	51.2	38.1	7.1	3.6		100.0

Table 3.7

		Conduct Expectation, Time 2					
		1	2	3	4	Total	Percent
Conduct Mark, Time 2	1	23	5	2		30	35.7
	2	23	10	4		37	44.0
	3	4	6	2	1	13	15.5
	4	1	2	1		4	4.8
	Total	51	23	9	1	84	
	Percent	60.7	27.4	10.7	1.2		100.0

Arithmetic

The marks given in arithmetic are considerably higher on the second report card--over 40% A's vs. 22% A's earlier. The combined number of B's and A's is about the same on the two occasions, however. There is noticeably less pessimism on the part of the children at the end of the year compared to midyear for 11% fear the lower two grades whereas earlier 21% harbored these fears. There has also been a reduction of extreme optimism--51% now look for A's compared to 61% earlier. For arithmetic the net effect is to produce a within-chance amount of correspondence between marks and expectations and departures about equally in the overly optimistic and overly pessimistic directions. The correspondence between expectations and marks in arithmetic on the second report card, in fact, looks almost random.

Conduct

For conduct there is likewise no significant matching between children's expectations and marks received (42% observed vs. 35% expected). Of more interest is the marked asymmetry in terms of marks failing to attain expectation level--44% of the children get a lower mark than they expect which implies that when expectations and marks disagree, there is a highly significant tendency for marks not to equal expectations ($\chi^2_1 = 12.76$, $p < .01$). (For the first report card matching was within chance expectancy and there was a slight but not significant tendency for marks to be less than expected.) Teachers are marking somewhat harder in conduct on the second report card than on the first--7% less A's and 5% more C's and D's.

In summary at both midyear and year end children's expectations in all three areas show only a chance level of matching with the marks they receive. There is a consistent pattern (in all areas at both midyear and year end) of children's expectations exceeding their marks if the two did not match but this only attains conventional significance levels for mid-year arithmetic, and year end reading and conduct.

Discrepances Between Expectations and Report Cards Over the Second Grade

Midyear, Second Grade

By the middle of second grade the child has received two second-grade report cards. Only the latter one (given at midyear) was examined.

Reading

As was true throughout first grade, in reading there is no significant matching between marks and children's expectation levels. As was true at midyear of first grade (but not at the end of first grade) there is no significant tendency toward over-optimism or over-pessimism in children whose hopes and marks differ in reading.

Table 3.8

		Reading Expectation, Time 3					
		1	2	3	4	Total	Percent
Reading Mark, Time 3	1	12	17	1	1	31	32.3
	2	18	22	6	1	47	49.0
	3	7	7	1		15	15.6
	4	2	1			3	3.1
	Total	39	47	8	2	96	
	Percent	40.6	49.0	8.3	2.1		100.0

Table 3.9

		Arithmetic Expectation, Time 3					
		1	2	3	4	Total	Percent
Arithmetic Mark, Time 3	1	9	5			14	14.4
	2	30	25	6	3	64	66.0
	3	8	4	1	5	18	18.6
	4		1			1	1.0
	Total	47	35	7	8	97	
	Percent	48.5	36.1	7.2	8.2		100.0

Table 3.10

		Conduct Expectation, Time 3					
		1	2	3	4	Total	Percent
Conduct Mark, Time 3	1	18	13	2		33	34.0
	2	19	25	7	2	53	54.6
	3	4	1	3	2	10	10.3
	4			1		1	1.0
	Total	41	39	13	4	97	
	Percent	42.3	40.2	13.4	4.1		100.0

Arithmetic

In arithmetic likewise there is no significant matching at the middle of grade 2, continuing the state of affairs noted throughout grade 1. There is a noticeable tendency for marks to be less than expected, however, for 44% of the children receive a mark lower than they had hoped for. This may be a consequence of a noticeably increased severity in marking. Whereas 94% of children had received B's and A's in arithmetic at the end of first grade, and 41% of these were A's, by the middle of second grade 80% are receiving B's and A's with only 14% of these A's. A large number of children (31%) expect an A but receive a B. This is the chief factor responsible for significant asymmetry in Table 3.9 whereby expectations exceed marks ($\chi^2_1 = 9.29$, $p < .01$). (This kind of asymmetry prevailed at midyear of grade 1 but not at the end of grade 1.)

Conduct

In conduct there is significant above chance matching (47% receive the mark they expect, $z = 2.00$, $p < .05$) and this is in contrast to lack of matching noted at both mid- and end-of-year in grade 1. Whereas at the end of grade one there was a significant tendency for the non-matching students to do worse than they expected, at the middle of second grade both positive and negative discrepances occur to about the same extent. Teachers give a few more B's and a few less C's and D's at the middle of second grade compared to the end of first grade.

End of Second Grade

Reading

At the end of second grade for the first time there is a highly significant match between expectations and marks in reading ($z = 2.86$, $p < .01$). The correspondence comprises 58% of the cases. Children's expectations are a little higher than they were at the middle of second grade, but the average mark awarded has also risen. About 92% of the children get A's or B's vs. 81% earlier.

Fewer of the children expect A's than was true at the end of first grade (46% compared to 62%), although the overall expectation for A's and B's is practically universal at both times (98% at the end of first grade and 94% at the end of second grade).

Arithmetic

There is likewise a significant match in marks expected and received in arithmetic at the end of second grade, for 57% receive exactly what they expect ($z = 3.72$, $p < .01$). Of the remainder a significant number (31% vs. 12%) ($\chi^2_1 = 8.40$, $p < .01$) do worse than they had hoped. Thus compared to the middle of second grade, children's ability to forecast has improved (there was no significant matching then or earlier in first grade). The errors in forecasting are still occurring on the optimistic

Table 3.11

		Reading Expectation, Time 4					
		1	2	3	4	Total	Percent
Reading Mark, Time 4	1	32	18			50	50.5
	2	13	24	4		41	41.4
	3		6	1	1	8	8.1
	Total	45	48	5	1	99	
	Percent	45.5	48.5	5.1	1.0		100.0

Table 3.12

		Arithmetic Expectation, Time 4					
		1	2	3	4	Total	Percent
Arithmetic Mark, Time 4	1	16	7			23	23.2
	2	24	34	3		61	61.6
	3	1	5	6	2	14	14.1
	4		1			1	1.0
	Total	41	47	9	2	99	
	Percent	41.4	47.5	9.1	2.0		100.0

Table 3.13

		Conduct Expectation, Time 4				
		1	2	3	Total	Percent
Conduct Mark, Time 4	1	24	17	1	42	42.4
	2	14	25	5	44	44.4
	3	2	7	2	11	11.1
	4	1		1	2	2.0
	Total	41	49	9	99	
	Percent	41.4	49.5	9.1		100.0

side--odds of about 10 to 4 for optimism, continuing a trend noted in the middle of second grade and the middle of first grade but not at the end of first grade.

The "over-optimism" may be more a function of different marking standards being applied than of changes in expectations, however. That is, children's forecasts may be more consistent than teachers' marking (or ultimately even children's performance). At the end of first grade only 6% of children receive less than a B in arithmetic whereas at the two subsequent semester points 20% and 15% receive less than a B. These changes in grading standards over time approximate the picture for reading. There about 16% get less than a B at the end of first grade, with 19% and 8% showing less than a B at the two second grade points.

Conduct

In conduct by the end of second grade there is a significant correspondence--52% matches--between expectations and marks ($z = 2.23$, $p < .05$). Of those whose marks do not match, over-optimists and over-pessimists are about equal in number (25% and 23% respectively). Teachers give 9% more A's at the end of second grade than in the middle. The overall division between high (A and B) and low (C and D) marks is fairly constant over the entire 2-year period. There are shifts, as noted, between the relative proportions of A's and B's.

In summary the second grade displays a marked change in the degree of matching between marks and children's expectations. Throughout grade one matching occurred at only chance levels but by the end of grade two all the marking areas displayed significantly above chance levels of matching. This indicates some fundamental change has transpired over the course of the second year. Some causal relationship has begun to display its effects. This suggests the end of grade one and all of grade two should provide fertile ground for beginning to sort out the causal relations linking the various variables of interest. The fact that the relationship developed in conduct before it did in the substantive academic areas may be a function of the child's differential monitoring and control capabilities regarding conduct.

Over the two year span only in arithmetic do children display a continual tendency to over-optimism regarding their mark. This occurs at 3 out of the 4 observation periods--the exception being time 2. Reading and conduct display no over-optimism in 3 of the 4 observation periods--the exception again being time 2. Changes in the mark distributions assigned (whether it be due to changes in the difficulty of subject matter, changes in teacher's marking severity, etc.) influence over- or under-optimism as much as do expectations themselves, so the degree of over- or under-optimism might be expected to be reasonably variable (as it is).

How Do Children's Expectations Change Over First Grade Year?*

Reading

If hope springs eternal in the human breast, it does so also in the very young breast for the marginal distribution of children's reading expectations at the end of the second semester of grade one is more skewed than the similar distribution earlier in the year. Whereas early in the year 12 children look for 3's or 4's, now only 3 children look for such low grades. Four of those formerly pessimistic 12 now even look for A's. In addition the lopsidedness between A's and B's at the end of the first semester (about twice as many expect A's as B's) diminishes very little despite the teachers' awarding of marks in reverse ratio. The first marks awarded include noticeably more B's than A's in about a 3:2 ratio. Children thus have unrealistically high expectations about A's at the start but the awarding of marks appears to do little to change these unrealistic expectations, so far at least. The awarding of marks (or possibly some other contemporaneous factor) does appear to dissipate the worst fears of the young pessimists, however. The exaggeration of the skewing of the expectation marginal is not quite enough to produce a significant asymmetry around the main diagonal ($\chi^2_1 = 3.56$, N.S.).

More than half (60%) of the children hold the same expectations at two points in time. This agreement is highly significant ($z = 2.77$, $p < .01$).

Arithmetic

There is less consistency over time in children's expectations for arithmetic, less than half (47%) showing perfect consistency. This amount of consistency is significant above chance levels, however ($z = 2.05$, $p < .05$). There is a moderate increase in optimism among those (24%) who expected to do the worst (3's and 4's). Interestingly ten of the 20 who earlier expected a C or a D expect an A at the end of the year. There are some marked counter movements as well--6% of the children display expectation declines of 2 or 3 units. Of the large number (59%) expecting 1's at the end of the first semester, over half are still looking for 1's at the end of the second semester. As with reading, there is no significant asymmetry ($\chi^2_1 = 0.36$, N.S.).

Conduct

Perhaps because many good marks in conduct were given, expectations in this area are generally moving in a more positive direction. There is little pattern in Table 3.16, however. Children's expectations at the two points in time do not match any better than would be expected by chance and there is no significant tendency toward increasing or declining expectations in those not matching. Of the 14 children expecting a C or a D at midyear, 6 or 43% come to expect an A at year end.

---

*Two report cards are issued in the second semester of grade one. Only the second report card--the one issued at the end of the semester (and hence also at the end of grade one)--is discussed here.

Table 3.14

		Reading Expectation, Time 2					
		1	2	3	4	Total	Percent
Reading Expectation, Time 1	1	37	10	1		48	56.5
	2	12	13			25	29.4
	3	1	4			5	5.9
	4	3	2	1	1	7	8.2
	Total	53	29	2	1	85	
	Percent	62.4	34.1	2.4	1.2		100.0

Table 3.15

		Arithmetic Expectation, Time 2					
		1	2	3	4	Total	Percent
Arithmetic Expectation, Time 1	1	28	18	2	2	50	58.8
	2	4	9	1	1	15	17.6
	3	8	5	3	1	17	20.0
	4	2		1		3	3.5
	Total	42	32	7	4	85	
	Percent	49.4	37.6	8.2	4.7		100.0

Table 3.16

		Conduct Expectation, Time 2					
		1	2	3	4	Total	Percent
Conduct Expectation, Time 1	1	30	12	3		45	52.9
	2	18	5	3		26	30.6
	3	3	3	2		8	9.4
	4	3	1	1	1	6	7.1
	Total	54	21	9	1	85	
	Percent	63.5	24.7	10.6	1.2		100.0

In summary, while expectations in the substantive areas of reading and arithmetic show significant stability from midyear to year end, conduct does not. In all three areas a substantial proportion (about 40%) of those expecting the two lowest marks (C's or D's) at midyear come to expect the highest mark possible (an A) by year end. The majority of those expecting C's or D's at midyear expect at least an average mark (B) by year end in all three areas. Expectations show no evidence of being permanently depressed if the child's first expectation was for failure.

How Do Children's Expectations Change from Grade 1 to Grade 2?

Reading

Well over half (60%) of the children have the same expectation for the end of the first semester, grade 2, as they held at the end of first grade. This is a highly significant match ($z = 2.51$, $p < .05$). Of those who shift, twice as many (27%) shift down as up (13%), but this does not attain conventional significance levels ($\chi_1^2 = 3.23$). The decline in the number of children expecting an A (50% vs. 62%) is what is producing the observed χ^2.

Arithmetic

The same amount of consistency between expectations at the end of first grade and second grade midyear expectations (59%) for arithmetic occurs as was noted for reading. This amount of matching is above chance levels (41%) and is highly significant ($z = 3.33$, $p < .01$). There are, however, almost equal numbers of children shifting up and down. Also the marginals at each time point are almost identical. Roughly twice as many children look for C's and D's in arithmetic as in reading, but the overall percentage (10%) is still small.

Conduct

There is significant matching in expectations for conduct at the end of first grade and midyear of second grade ($z = 2.45$, $p < .05$). The teacher's bearing down on conduct grades at the end of grade 1 may have led to a decrease in children's optimism about conduct grades near the middle of grade 2. Whereas at the end of first grade, about 63% looked for an A, 42% now expect an A. Four more children expect C's than formerly, and 2 more expect D's. There is a significant overall downward shift in expectations for conduct grades ($\chi_1^2 = 10.53$) over this time interval (12% move up and 37% move down).

Table 3.17

		Child's Reading Expectation, Time 3				
		1	2	3	Total	Percent
Child's Reading Expectation, Time 2	1	30	17	1	48	61.5
	2	7	17	3	27	34.6
	3	1	1		2	2.6
	4	1			1	1.3
	Total	39	35	4	78	
	Percent	50.0	44.9	5.1		100.0

Table 3.18

		Child's Arithmetic Expectation, Time 3					
		1	2	3	4	Total	Percent
Child's Arithmetic Expectation, Time 2	1	27	11		2	40	51.3
	2	13	15	1		29	37.2
	3	1	2	1	1	5	6.4
	4		1		3	4	5.1
	Total	41	29	2	6	78	
	Percent	52.6	37.2	2.6	7.7		100.0

Table 3.19

		Child's Conduct Expectation, Time 3					
		1	2	3	4	Total	Percent
Child's Conduct Expectation, Time 2	1	26	19	4		49	62.8
	2	6	10	3	2	21	26.9
	3		2	4	1	7	9.0
	4	1				1	1.3
	Total	33	31	11	3	78	
	Percent	42.3	39.7	14.1	3.8		100.0

How Do Children's Expectations Change Over the Second Grade Year?

Reading

There is some mild increase in children's hopes about reading over the second grade year as seen in Table 3.20--whereas 10% look for C's and D's in the middle of the year, that percentage has shrunk to about 6% by the end of the year. The shifts upward (20%) do not differ significantly ($\chi_1 = 0.82$, N.S.) from those downward (27%).

There is significant consistency in reading expectations over the second grade year, with 53% matching between midyear and end-of-year second grade ($z = 1.95$, $p = .05$). The consistency over second grade however is less, although not significantly so, than that noted over the first grade year (60%), or between the end of first grade and the middle of second (60%).

Perhaps the most interesting thing is the change between the earliest expectations and later ones in terms of extreme pessimism. At first sampling 8% of the children expected the lowest grade. As time goes on, practically no one ever again has such low hopes (the values ranging from 1 to 2%). Another rather surprising fact is the fluctuation from one time to the next for a few children. One would think that if a child's expectations did not remain stable, they would probably shift up a notch or down a notch. From the middle to the end of first grade 8% of children shift two or more steps, and from the middle to the end of second grade 7% of the children do so. These percentages do not look large but with a sample of around 100, they imply that about 2 children in every classroom have extremely labile expectations even within a single year.

Arithmetic

In arithmetic the consistency in expectations (46%) over the second grade does not significantly exceed chance (39%) ($z = 1.43$, N.S.). This contrasts with a strong consistency (59%, $p < .01$) between the end of first grade and the middle of second, and a lesser but still significant consistency over the first grade year (47%, $p < .05$).

There are shifts upward and downward in about equal numbers (28% and 27% respectively).

Conduct

In conduct the consistency over the second grade year (48%) does not significantly exceed chance (39%) although it is not far from the 5% level ($z = 1.87$). There is no significant asymmetry up or down in Table 3.25. This is the same state of affairs as that prevailing over the first grade year. Surprisingly, in view of this lack of consistency within both years, there is some consistency between the end of first grade and the middle of second (51%, $z = 2.45$, $p < .05$). This was accompanied by a significant

Table 3.20

		Reading Expectation, Time 4					
		1	2	3	4	Total	Percent
Reading Expectation, Time 3	1	22	16	1		39	41.5
	2	18	26	1	1	46	48.9
	3	3	2	2		7	7.4
	4	1	1			2	2.1
	Total	44	45	4	1	94	
	Percent	46.8	47.9	4.3	1.1		100.0

Table 3.21

		Arithmetic Expectation, Time 4					
		1	2	3	4	Total	Percent
Arithmetic Expectation, Time 3	1	23	19	3	1	46	48.9
	2	14	18	2		34	36.2
	3	1	3	2		6	6.4
	4	1	6	1		8	8.5
	Total	39	46	8	1	94	
	Percent	41.5	48.9	8.5	1.1		100.0

Table 3.22

		Conduct Expectation, Time 4				
		1	2	3	Total	Percent
Conduct Expectation, Time 3	1	20	16	3	39	41.5
	2	11	24	3	38	40.4
	3	4	8	1	13	13.8
	4	2	1	1	4	4.3
	Total	37	49	8	94	
	Percent	39.4	52.1	8.5		100.0

downward adjustment in expectations (12% moved up and 37% moved down, $\chi_1^2 = 10.52$, $p < .01$). Most of the down shift was accounted for by a reduction from expecting A's to B's.

In summary, if we interpret consistency or above chance levels of matching between expectation over a time interval as being indicative of some causal connection between the expectations at the two times (either as the earlier being a direct cause of the later, or both being the result of a stable third factor such as ability or parental expectations) we see the data on levels of significance for both grades 1 and 2 provide no simple answer regarding such causal relations. The pattern of the significance of matching of expectations from one time to the next is as follows:

	T1-2	T2-3	T3-4
Reading	.01	.05	.05
Arithmetic	.05	.01	N.S.
Conduct	N.S.	.05	N.S.

All the non-significant cases display deviations in the direction of above-chance matching and almost reaching significance in the case of T3-4 conduct. (The chance expectations are exceeded by the following percentage points for reading, arithmetic and conduct respectively: T1-2, 14, 9, 3; T2-3, 14, 18, 12; T3-4, 10, 7, 9.) The consistency of the above chance matching, nine cases out of nine being in one direction ($p < .01$ on a binomial test), suggests some causal relation does exist.

It should be noted that children with low expectations do not provide an overly abundant input of matched expectation cases. That is, those with low expectations do not show exceedingly strong tendencies for holding the same low expectation over time. In general the majority of those expecting C's and D's at any initial time, hold expectations for average or top marks by the second time. (T2-3 arithmetic and conduct expectations are exceptions where 56% and 63% respectively maintain low expectations over time.)

Those children whose expectations did not match at the consecutive times, in general, show no tendency to either lower or raise their expectations. Conduct expectations between end of grade one and middle of grade two are an exception. Here the asymmetry was toward a decline in expectation.

How Expectations Change By Expectation Level
(Omitted)

See Chapter 5 for first year results.

How Marks Compare in Different Areas

Results from the comparison of marks in different areas are most meaningful when looked at over the whole time span available (two years) and over all possible combinations of mark pairs. Two questions arise: (1) How consistent are children's marks in any two areas (a matching problem)? (2) For the cases that do not match, do children's marks in one area generally exceed marks in the other area (an asymmetry problem)?

Consistency or Matching

The significance of the matching between marks in the different areas is summarized in the following table.

	T1 (N = 85)	T2 (N = 86)	T3 (N = 102)	T4 (N = 103)
Reading-Arithmetic	.01	.01	.01	.05
Arithmetic-Conduct	.05	N.S.	N.S.	N.S.
Reading-Conduct	N.S.	N.S. (z = 1.84)	.01	.01

The picture is quite clear. The academic areas (reading and arithmetic) show a consistent above chance level of matching. Marks in conduct, the behavioral area, generally show no significant matching with arithmetic mark (T1 being an exception where a weak and barely significant (z = 1.98) relation occurs). Conduct marks, on the other hand, show a trend toward becoming consistent with reading marks.

The level of chance matching between marks in any two areas is generally near 39% while the significantly above chance matches range from 49% to 64%. Considerable variation between children's marks in any two areas remains even when above chance levels of matching are present.

When above chance matching occurs it seems to occur at all mark levels. There is no apparent tendency for those with high, average or low marks to differentially account for excess matches. Where excess matching exists children with high, average and low marks all seem to have contributed to the excess.

Asymmetry

The following table presents the pattern of significance associated with asymmetry among unmatched marks. In this table R > A is an abreviated notation indicating the asymmetry is of the form where reading marks are higher than arithmetic marks.

	T1	T2	T3	T4
Reading-Arithmetic	N.S.	N.S.	.05(R > A)	.01(R > A)
Arithmetic-Conduct	N.S.	.05(A > C)	.01(C > A)	.01(C > A)
Reading-Conduct	N.S.	N.S.	N.S.	N.S.

The four significant values in the upper right corner of the table arise because of the relatively hard marking of arithmetic at both times during the second year. This means that in comparisons of reading and conduct with arithmetic the unmatched cases tend to show the arithmetic mark as being the lower. The remaining significant value is a product of fairly easy arithmetic marking and fairly hard conduct marking at the end of the second year. Neither of these by themselves, however, are severe enough to produce a significant relationship with reading which is moderately marked.

How Consistent are the Children's Marks over Time?

For the middle-class school there are three time periods for which mark changes are considered, namely, changes during the first year, between the end of grade one and the middle of grade two, and during the second year. The significance of the matching between marks over the time intervals is as follows.

	T1-2 (N = 78)	T2-3 (N = 82)	T3-4 (N = 103)
Reading	.01	N.S.	.01
Arithmetic	.01	N.S.	.01
Conduct	.01	N.S.	.01

The pattern is both obvious and somewhat startling. In all three areas children's marks <u>within</u> both the first and the second grades show significant above chance levels of consistency or matching. The chance levels of matching range from 36% to 49% while the observed significant matches ranged from 55% to 73%. (The observed insignificant values ranged from 41% to 45% matches.) Clearly, within each grade the marks the children receive are quite consistent. About two-thirds of the children receive identical marks at the two observations in each year.

The matching of marks between the end of grade one and the middle of grade two shows a completely different picture. Here, none of the areas shows a significantly above chance consistency.

Substantively there are two styles of explanation for this finding. First, because in this school each child's performance is marked

according to his own ability (their marks approximate the children's efforts) the finding can be restated as follows: Any given teacher is quite consistent in his/her evaluation of a child's scholastic efforts but teachers differ with each other in their evaluation of effort. The second explanation is that of a year X child interaction. Each school year can be envisaged as a set of environmental variables for each child that are largely consistent throughout a year yet which change between years (e.g., teacher personalities, presence of a child's best friend in his class, etc.). If children's efforts interact with such variables (different environments stimulating or depressing different children) the observed pattern of mark consistency would be expected. The above alternatives reflect one's choice of a model that postulates variation in teacher's marking (with children expending stable efforts) or variation in children's efforts (with reliable inter-teacher marking). The real world probably evidences a miniture of these "ideal types".

For the cases that do not display matching of marks at the comparison times, it can be asked if either high or low marks tend to predominate. The following table summarizes the significance of asymmetry in the comparison of marks at the various times as tested by X_1^2. (Use of 2 > 1 etc. in this table indicates marks at time 2 are higher than marks at time 1, etc.)

	T1-2	T2-3	T3-4
Reading	N.S.	N.S.	.01(4>3)
Arithmetic	.01(2>1)	.01(2>3)	.01(4>3)
Conduct	N.S.	N.S.	N.S.

Conduct marks remain constant (on the average) throughout the first two years. Arithmetic marks increase over the first year. Indeed they are very high at year end--41% get A's, the highest mark possible. The drop in arithmetic marks from the end of grade one to mid-grade two is more than "getting back to a reasonable mark distribution" for the arithmetic marks at the later time have the lowest average mark ever assigned in any of the areas (2.07). When arithmetic marks again rise over the second year they do not rise above the level attained the first year. The rise in conduct marks over the second year to the highest average observed (1.60) places the next year's teacher in a difficult "personnel management" position. This teacher is forced to mark comparitively hard to produce a reasonable mark distribution and extremely hard comparatively if he/she wants the children's marks to be able to "show come improvement" over the third grade year. A child's marks cannot improve if they are already at the top--even if they got to the top by the last year's teacher's lenient year-end marking.

In all areas teachers generally assign few year end marks that differ by more than one point from the midyear mark assigned. The percentages are 1%, 1% and 3% for reading, arithmetic and conduct respectively in grade one and 2%, 1% and 1% respectively in grade two. Between years one and two (times 2 and 3) there are a few more changes of two or more points in marks (5%, 4% and 4% respectively) but these are still quite small values.

How Expectations Compare in Different Areas

At End of First Grade

There is a surprising negative association between expectations for marks in reading and marks in arithmetic at the end of first grade. Only 32% gave the same expectations for the two areas, a failure of matching that is highly significant ($z = -2.61$, $p < .01$). (An average of 45% matching would be expected by chance.) This is contrary to the matching in same-subject expectations over time. For instance there is significant matching for first graders between first and second semester expectations in reading (60%, $p < .01$), and in arithmetic (47%, $p < .05$), although not in conduct.

There is a degree of matching that would be expected by chance between reading and conduct expectations, and between arithmetic and conduct expectations at the end of first grade. None of the comparisons of expectations between areas show expectations in one area to significantly exceed expectation in another.

At End of Second Grade

There is no significant matching between children's expectations for reading and for arithmetic at the end of second grade (41% match). (This lack of association is to be contrasted with the significant negative association at the end of first grade.) There, furthermore, is no significant tendency for expectations in one area to be higher than the other.

There is also only the degree of matching that would be predicted by chance between expectations in arithmetic and conduct and between those in reading and conduct. For neither of these comparisons is there a tendency for children whose expectations do not match to have expectations higher in one area.

It is of interest to note that while a large majority of expectations in different areas fall within one unit of each other, there are a reasonable number of cases where this is not true. For reading and arithmetic, arithmetic and conduct, and reading and conduct the respective percentages of expectations differing by two or more points are 8%, 10% and 11% at the end of first grade, and 5%, 7%, and 5% at the end of second grade. Very few children (an average of about 1.5%) hold low expectations (i.e., for C's or D's) in any two areas. None hold low expectations in all three areas at the end of grade two. This suggests that in the middle-class school at least, if ameliorative intervention via raising expectations was to be engaged in it should be subject and child intervention as opposed to simply child oriented intervention. One would not attempt to "raise the child's expectations" but rather "raise his/her reading or arithmetic or conduct expectation", as the other expectations are usually already at average or above average levels.

What is Impact of First Mark on Expectation Later?

Reading

There is a surprising lack of consistency between what the child expects at the end of grade 1 with his mark earlier in the year. Less than 50% of the children expect again what they got earlier, and the amount of matching does not exceed what would be expected by chance. For the cases that do not match (51%) almost all deviate in an optimistic direction. Thus almost half of those expecting a 1 got a 2 last time. About 10% of the students look for a decrease, all but one are in the group who expect a 2 this time but who received a 1 last time. Fully 41% look for a higher grade. (This asymmetry is highly significant; $X_1 = 16.10$, $p < .01$). If the teachers were to conform to students' hopes they would have to award 63% of the students A's! (They actually had given over one-third A's earlier and do give 47% A's at year end.)

Arithmetic

There is even less correspondence between the child's hopes at the end of the year for arithmetic and his earlier mark than was true for reading. The correspondence between the first mark in arithmetic and hopes at the end of first grade does not exceed an amount predicted by chance. Also, the asymmetry is likewise in an optimistic direction, although here not quite statistically significant ($X_1^2 = 3.45$). An interesting fact about this relationship is that some children (5%) show a correspondence between low marks and hopes for high marks (1's and 2's). Conversely eleven children (13%) got 2's earlier but expect to get 3's and 4's at the end of the year. The teachers actually had not given anyone a "4" in either arithmetic or reading and children do not look for 4's in reading or conduct. Arithmetic seems to be a subject which provokes more anxiety (but this could be a manifestation of a small case base). Some children's fears and other's high hopes are not grounded in reality.

Conduct

There is a significant amount of matching ($z = 4.16$, $p < .01$)--56%--between what the child expects for his mark in conduct and the mark he received first in this subject. There is also a significant ($X_1^2 = 13.44$, $p < .01$) tendency toward optimism in those whose expectations depart from the earlier mark, for 81% of those expecting their mark to change, expect it will rise.

What is the Impact of Later Marks on Expectations?

Reading

The degree of matching between the mark in reading at the end of first grade and the mark expected in the middle of second grade is again within chance expectancy. About 47% expect to receive what they received earlier,

Table 3.23

		Child's Reading Expectation, Time 2				
		1	2	3	Total	Percent
Reading Mark, Time 1	1	22	7		29	35.4
	2	25	17	1	43	52.4
	3	5	4	1	10	12.2
	Total	52	28	2	82	
	Percent	63.4	34.1	2.4		100.0

Table 3.24

		Child's Arithmetic Expectation, Time 2					
		1	2	3	4	Total	Percent
Arithmetic Mark, Time 1	1	11	7			18	22.0
	2	27	22	7	4	60	73.2
	3	1	3			4	4.9
	Total	39	32	7	4	82	
	Percent	47.6	39.0	8.5	4.9		100.0

Table 3.25

		Child's Conduct Expectation, Time 2				
		1	2	3	Total	Percent
Conduct Mark, Time 1	1	30	2	2	34	41.5
	2	18	13	3	34	41.5
	3	4	6	3	13	15.9
	4		1		1	1.2
	Total	52	22	8	82	
	Percent	63.4	26.8	9.8		100.0

and the departures in optimistic or pessimistic directions are not very different (30% and 23% respectively). (After the first mark there was a similar lack of consistency between expectations at the end of the year and midyear marks, but hopes were notably in an optimistic direction.)

The correspondence between marks in reading in the middle of the second year and expectations at the end of that year still does not exceed what would be predicted by chance, given the marginal distributions. But now there is a resurgence of optimism. Significantly more (37%) have expectations higher rather than lower (14%) than their marks ($\chi_1^2 = 9.68$, $p < .01$). It thus appears that the early optimism, which appeared to subside, surfaces again in second grade.

Arithmetic

In arithmetic there is likewise no significant matching between the mark at the end of the first grade and the expectation held in the midyear following. There is some tendency toward optimism, but it is not significant. This replicates the picture seen for the relation between first grade midyear marks in arithmetic and expectations later that year.

Again there is failure to match significantly from midyear to the end of second grade in arithmetic. There is, however, a highly significant ($\chi_1^2 = 18.96$, $p < .01$) tendency for expectations to be higher than the mark received earlier, for 43% expect a higher mark whereas 11% expect a lower mark. Although only 15% of children earlier had received a mark of "A", 41% of the children look for that mark at the end of the year.

Conduct

In conduct there is not a significant degree of matching between the mark received at the end of first grade and the expectation at the middle of second grade, although there was significant matching of this sort from the middle to the end of first grade. Also the scatter above and below is now similar whereas earlier expectations were significantly optimistic.

Between the middle and the end of second grade, significant matching again appears (58%) ($z = 3.75$, $p < .01$). Also optimism does exceed pessimism to some extent but not significantly.

It is tempting to speculate that "conduct" is a teacher-specific activity and one which students feel they can effectively alter. Students may have to learn what a particular teacher regards as acceptable and unacceptable conduct, and they may expect little carry-over between first and second grade. The students' expectation of little carry-over between years is substantiated by the insignificant matching for conduct between times 2 and 3 but in this area students' expectations come to match the mark given at midyear (and this happens both years) which suggests students feel they cannot effectively alter their conduct. They adopt their previous mark as their present expectation. The data contradict the above "tempting speculation".

To sum up: The pattern of the results for all the comparisons of expectations at times subsequent to mark assignments are presented in the following table. (M = matching, A = asymmetry and again E>M means for the asymmetry expectations are higher than earlier mark. The marks are always observed at the earlier time period--expectations at the later time period.)

	Times 1 and 2		Times 2 and 3		Times 3 and 4	
	M	A	M	A	M	A
	(N = 82)		(N = 77)		(N = 99)	
Reading	N.S.	.01(E>M)	N.S.	N.S.	N.S.	.01(E>M)
Arithmetic	N.S.	N.S.	N.S.	N.S.	N.S.	.01(E>M)
Conduct	.01	.01(E>M)	N.S.	N.S.	N.S.	N.S.

In reading and arithmetic children show no sign of adopting an earlier mark and maintaining that specific mark as their later expectation. Surprisingly, in conduct this does occur within grades one and two (but not between grades). This is directly opposite to what would be expected if children felt they could effectively alter their conduct marks during grades one and two!

Those dedicated to an "adopting of marks as later expectations" theory might suggest that the issuing of a report card between each of our "earlier mark" "later expectation" pairs renders this an insufficient test of the theory, the claim being that children may be showing an extreme form of "adopting marks" at each successive mark assignment. This suggests that only the previous mark assignment can be used to test the theory, not a mark assignment one removed from the time of the measurement of expectations. This claim has little foundation, however. It has already been reported in the discussion of mark consistency over time that marks show highly significant and substantial consistency within each grade. Therefore a child's mark that appears between the time of the initial mark determination and the time of the expectation determination is very likely to be identical with the mark utilized in the comparison, thus removing the objection. (The case between years is not as clear--here no significant matching of marks occurred (1/3 to 1/2 of the marks still matched by chance) but this was not a matching between end of grade one and the first marks of grade two, as would be necessary to justifiably make this argument.)

The asymmetry that occurs is only present within grades one and two (not between the grades) and is always of the type where present expectations exceed earlier mark. This is not surprising since expectations generally exceeded marks at all times. In particular the arithmetic asymmetry for times 3 and 4 is partially an artifact of the extremely hard arithmetic marking at the middle of grade two (time 3). This asymmetry would be expected if the children simply retained their previous expectations. There seems no basis on which to predict when expectations will exceed previous marks, other than that such an excess appears within

grades. While asymmetry in this type of table is a sign of unrealism if expectations drastically exceed previous marks, it can also be a sign of realistic striving if expectations moderately exceed previous marks. (The latter is probably the case for 87% of those whose expectations exceed previous marks in the tables with significant asymmetry, since their expectations exceed their mark by only one point.)

Effects of Feedback on Marks and Expectations

Discussions of expectation-mark agreement in terms of marks received (midyear and end of year) and in terms of expectation level are found in Chapter 5 as is a discussion of feedback effects for both middle-class cohorts in first grade. In this section we will review briefly the findings on feedback for Cohort 1 as a prelude to discussion of feedback effects in second grade for that cohort.

"Feedback" in this discussion will be called "positive," "neutral," or "negative" depending on whether the first mark received was higher than expected (positive), the same, or less than expected (negative).

In reading there was a strong tendency for first grade children to persist in the same expectation at year's end if midyear marks matched expectations <u>or</u> if marks were less than expected. If marks exceeded expectations most (79%) revised hopes upward. In arithmetic and conduct there was a similar (but insignificant) tendency except that children were less apt to persist in expectations if feedback was negative. For marks in both reading and arithmetic most of those whose marks rise were children who expected to do better than they actually did at midyear.

Second Grade Marks

Reading

Table 3.26 presents the changes in marks that occurred between the middle of grade two (time 3) and the end of grade two (time 4) following the various possible combinations of marks and expectations that occurred at the middle of grade two. The minor diagonal of this table is significantly ($z = 3.38$, $p < .01$) over-represented which indicates an above chance occurrence of cases where: marks move up if the child had earlier expected to do better, marks move down if the child had earlier expected to do worse, and marks remain the same if the child earlier got exactly what he expected. Reading marks respond more vigorously over the second grade to a positive discrepance (expectation exceeds mark) than to a negative discrepance. Sixty per cent of the cases that showed an initial positive discrepance had marks that moved up in response, while only 8% of the cases that showed an initial negative discrepance had marks that declined in response. (This is still above the chance level of 3% expected declines.)

Table 3.26

		Reading Mark Went U, S, D from Time 3 to Time 4			
		Up	Same	Down	
Child Did B, S, W than He Expected to Do in Reading, Time 3	Better	2	22	2	26
	Same	8	26	1	35
	Worse	21	14	0	35
		31	62	3	96

Table 3.27

		Arithmetic Mark Went U, S, D from Time 3 to Time 4			
		Up	Same	Down	
Child Did B, S, W than He Expected to Do in Arithmetic, Time 3	Better	3	15	1	19
	Same	4	29	2	35
	Worse	13	26	4	43
		20	70	7	97

Table 3.28

		Conduct Mark Went U, S, D from Time 3 to Time 4			
		Up	Same	Down	
Child Did B, S, W than He Expected to Do in Conduct, Time 3	Better	3	19	4	26
	Same	7	32	7	46
	Worse	11	12	2	25
		21	63	13	97

Arithmetic

Table 3.27 indicates that arithmetic shows a similar result to that in reading with some minor differences. The minor diagonal matching significantly exceeds chance expectation ($z = 1.80$, $p < .05$) using a one-tailed test but the over-representation characterizes only two of the three diagonal cells these show a) a mark rise following an initial positive expectation discrepance or b) no change in marks following no initial discrepance. Too few cases involve the falling of marks for any conclusions to be drawn. The movement of marks upward following a prior condition in which expectations exceeded marks, is not as strong in arithmetic as in reading (30% move up here while 60% move up in reading).

Conduct

A similar pattern prevails for conduct marks (Table 3.28). The minor diagonal is significantly over-represented ($z = 1.88$, $p < .05$) using a one-tailed test. The cells indicating downward mark movement following a negative discrepance and no mark change following no discrepance are only slightly over-represented (15% observed vs. 13% expected; and 70% observed vs. 65% expected respectively).

To sum up, in all three areas mark changes over the second year tend to follow expectations with the direction of movement predicted by the expectation aspect of the initial discrepance between marks and expectations. Marks move up more easily than down in response to an original discrepance, thus continuing the "mark" half of the bouyancy effect (both marks and expectations tending to move up to close an initial discrepance more easily than they move down to close an initial discrepance) noted in grade one.

Second Grade Expectations

Tables 3.29, 3.30, and 3.31 show the changes in expectations that occur during the second grade following the various possible types of mark-expectation discrepances, for reading, arithmetic and conduct respectively. The similarity of these tables allows their simultaneous discussion. In this section, values presented in triplicate are the respective values for reading, arithemtic, and conduct.

Significant above chance major diagonal matching occurs ($z = 5.70$, $p < .01$; $z = 4.82$, $p < .01$; $z = 6.33$, $p < .01$). This indicates that during the second grade children's expectations tend to: move up if they earlier had received a better mark than they expected, move down if they earlier had received a mark worse than they expected, and stay the same if they earlier received exactly what they expected. The rise in expectations following the receipt of a better mark than expected (69%, 78%, and 73%) is more pronounced than the fall in expectations following the receipt of a worse mark than expected (47%, 49%, and 60%). Conduct is somewhat different than the substantive areas in that negative feedback produces downward expectation revision with nearly the same certainty as positive feedback produces upward revision of expectations.

Table 3.29

		Reading Expectation Went U, S, D from Time 3 to Time 4			
		Up	Same	Down	
Child Did B, S, W than He Expected to Do in Reading, Time 3	Better	18	8	0	26
	Same	7	22	3	32
	Worse	0	18	16	34
		25	48	19	92

Table 3.30

		Arithmetic Expectation Went U, S, D from Time 3 to Time 4			
		Up	Same	Down	
Child Did B, S, W than He Expected to Do in Arithmetic, Time 3	Better	14	4	0	18
	Same	11	19	4	34
	Worse	1	20	20	41
		26	43	24	93

Table 3.31

		Conduct Expectation Went U, S, D from Time 3 to Time 4			
		Up	Same	Down	
Child Did B, S, W than He Expected to Do in Conduct, Time 3	Better	19	7	0	26
	Same	7	28	7	42
	Worse	1	9	15	25
		27	44	22	93

Expectations thus also continue to contribute to a bouyancy effect as noted for grade one. Expectations tend to rise to close a discrepance more readily than they fall to close a discrepance. (Note marks exceed expectations by almost exactly the same amount as expectations exceed marks when discrepances appear. This eliminates differing degrees of discrepance as an explanation.) The bouyancy effect (marks and expectations moving up to close an initial discrepance more readily than they move down to close an initial discrepance), then, continues on into grade two.

This section on expectation changes and the preceeding one on mark changes provide evidence for the causal efficacy of marks on expectations and the causal efficacy of expectations on marks in the second grade, thus extending this observation beyond the first grade as reported in other sections of this report.

It should be noted that both marks and expectations are reasonably stable during grade two. Marks remain the same at both times 3 and 4 in 65%, 72%, and 65% of the cases for reading, arithmetic and conduct respectively. Expectations are a little more volatile, yet they remain the same in 52%, 46%, and 47% of the cases for the respective areas.

Feedback Effects Between Grades One and Two

Having shown that expectations and marks both display causal efficacy within grades one and two, it can be asked whether the same holds true between grades one and two. The comparison involves effects of mark-expectation discrepances at the end of grade one (time 2) on marks and expectations at the middle of grade two (time 3).

For the children there have been substantial changes during this period, including summer vacation when they may have forgotten about their academic ambitions or fears. They have entered a new classroom, with some new student peers, a new teacher and more advanced subject matter. Also they have received one further report card during this period. There are thus many factors that could intervene and disrupt the causal effects observed within grades.

Tables 3.32, 3.33 and 3.34 show changes in marks, while Tables 3.35, 3.36, and 3.37 show changes in expectations, following year end mark-expectation discrepances, for reading, arithmetic and conduct respectively.

The following table presents the results for the tests of whether the diagonals (main diagonal in the case of expectation; minor diagonal in the case of marks) contain an over-representation of cases. Since for our purposes a change in a variable following a previous discrepance is evidence for causal efficacy these are really tests to see if the causal relations hold between years. (Column headings are dependent variables.)

Table 3.32

		Reading Mark Went U, S, D from Time 2 to Time 3			
		Up	Same	Down	
Child Did B, S, W than He Expected to Do in Reading, Time 2	Better	0	6	6	12
	Same	4	18	13	35
	Worse	14	10	8	32
		18	34	27	79

Table 3.33

		Arithmetic Mark Went U, S, D from Time 2 to Time 3			
		Up	Same	Down	
Child Did B, S, W than He Expected to Do in Arithmetic, Time 2	Better	2	6	13	21
	Same	4	15	17	36
	Worse	4	12	7	23
		10	33	37	80

Table 3.34

		Conduct Mark Went U, S, D from Time 2 to Time 3			
		Up	Same	Down	
Child Did B, S, W than He Expected to Do in Conduct, Time 2	Better	2	3	7	12
	Same	3	20	12	35
	Worse	20	12	1	33
		25	35	20	80

Table 3.35

		Reading Expectation Went U, S, D from Time 2 to Time 3			
		Up	Same	Down	
Child Did B, S, W than He Expected to Do in Reading, Time 2	Better	4	7	0	11
	Same	2	21	11	34
	Worse	3	18	10	31
		9	46	21	76

Table 3.36

		Arithmetic Expectation Went U, S, D from Time 2 to Time 3			
		Up	Same	Down	
Child Did B, S, W than He Expected to Do in Arithmetic, Time 2	Better	10	8	2	20
	Same	7	24	3	34
	Worse	0	13	9	22
		17	45	14	76

Table 3.37

		Conduct Expectation Went U, S, D from Time 2 to Time 3			
		Up	Same	Down	
Child Did B, S, W than He Expected to Do in Conduct, Time 2	Better	5	5	1	11
	Same	2	20	11	33
	Worse	2	14	16	32
		9	39	28	76

Significance of Diagonal (Main or Minor) Matching

	Changes in Marks	Changes in Expectations
Reading	.01 (z = 2.78)	N.S. (z = 1.05)
Arithmetic	N.S. (z = 1.01)	.01 (z = 3.62)
Conduct	.01 (z = 4.33)	.01 (z = 2.58)

Clearly some causal efficacy for both marks on expectations and expectations on marks is present.

Reading and conduct marks respond to year end discrepances as do arithmetic and conduct expectations.

While arithmetic marks respond somewhat to expectations they do not respond significantly. This might be a function of changed subject matter, but probably reflects the ambiguity of the evaluations of performance in the two areas. Arithmetic is more easily judged "right" and "wrong" and a child's performance is less affected by sheer effort. Reading performance depends in part on intonation, "expression", and interpersonal variables that are harder to judge reliably. A teacher's judgments may be more easily shaded in reading or conduct and a child evidencing high self-expectations in these two areas may exert considerable "pull" on a teacher for a better mark. The changes in expectations may be the fall out which occurs subsequent to a considerable rise in average marks in arithmetic over the first grade year.

It is worth noting that <u>all</u> diagonal cells in Tables 3.32 to 3.37 are over-represented, even when over-representation does not attain significance. This foreshadows the observation that the bouyancy effect (marks and expectations moving up to close an initial discrepance more readily than they move down to close an initial discrepance) does <u>not</u> operate between years. For changes in marks, downward movement following a negative discrepance (marks exceed expectations) is, if anything, more likely than an upward mark movement following a positive discrepance. For reading, arithmetic and conduct respectively the percentages of marks moving down (after a negative discrepance) are 50%, 62%, and 58%, and the percentages moving up (after a positive discrepance) are 44%, 17%, and 61%. The expectation aspect of the bouyancy effect has also largely disappeared between grades one and two. Expectations are about as likely to move down following a positive discrepance (expectation exceeds mark) as they are to move up following a negative discrepance (mark exceeds expectation). For reading, arithmetic and conduct respectively the percentages of expectations moving down following a positive discrepance (expectation exceeds mark) are 32%, 41% and 50%, and the percentages moving up after a negative discrepance (mark exceeds expectation) are 36%, 50%, and 45%.

<u>In sum</u> between grades one and two both marks and expectations in general seem to maintain their causal efficacy for one another when earlier

discrepances are compared with later mark changes or expectation changes. Two cases (reading expectations and arithmetic marks as dependent variables) show no significant effect but the deviations that do occur even in these cases are in the predicted directions. The buoyancy effect (especially the mark aspect of this effect) does not seem to operate between grades one and two.

Parents' Expectations

Compared to Children's Expectations

Reading

What is the degree of congruence between the expectations of parents and children in reading? In only 25 out of 80 cases (31%) for reading is there perfect agreement. A random shuffle would, on the average, produce 31% agreement so the amount of agreement is not statistically significant.

In 42 of 55 cases of disagreement (76%) the parent's expectation is lower. All of these instances represent a child expecting an A and the parent expecting a B or C. There are significantly fewer instances ($\chi_1^2 = 15.29$, $p < .01$) where children's expectations are lower than parents', 24% of the inconsistencies are of this type.

Another way to summarize the lack of congruence between the two sets of expectations (in addition to different means, standard deviations, margins, and only 31% agreement) is to compute a product-moment correlation as an index of the degree to which parents' and children's expectations covary. This correlation is practically zero (0.02). The margins of Table 3.38 show that approximately 90% of both parents and children expect an A or a B but that only about one-third of these expectations are consistent.

Arithmetic

The picture for arithmetic is in general similar to that for reading but there is a somewhat different kind of disagreement between parents and children. Again the actual amount of agreement is exactly what would be predicted by chance (30%). Again, also, many children (60%) expect the highest mark. For arithmetic, however, parents are more optimistic than they were for reading--28% of them (compared to 16% for reading) expect the highest grade.

The other striking difference between reading and arithmetic is that children are noticeably more pessimistic, 21% of them looking for a low grade (3 or 4). Parents are no more pessimistic for arithmetic than for reading (about 11% expect a 3, none expect a 4). The difference between parents' mean expectation and children's is 0.17 in arithmetic, about

Table 3.38

Parent's Reading Expectation, Time 1

		1	2	3	4	Total	Percent
Child's Reading Expectation, Time 1	1	6	35	7		48	60.0
	2	6	17			23	28.7
	3		2	1		3	3.7
	4	1	4		1	6	7.5
	Total	13	58	8	1	80	
	Percent	16.2	72.5	10.0	1.2		100.0

Table 3.39

Parent's Arithmetic Expectation, Time 1

		1	2	3	Total	Percent
Child's Arithmetic Expectation, Time 1	1	15	28	5	48	60.0
	2	4	8	3	15	18.7
	3	2	11	1	14	17.5
	4	1	2		3	3.7
	Total	22	49	9	80	
	Percent	27.5	61.2	11.2		100.0

Table 3.40

		Parent's Conduct Expectation, Time 1				
		1	2	3	Total	Percent
Child's Conduct Expectation, Time 1	1	9	26	6	41	54.7
	2	4	18	2	24	32.0
	3	2	3		5	6.7
	4		5		5	6.7
	Total	15	52	8	75	
	Percent	20.0	69.3	10.7		100.0

half the size of the mean difference for reading (0.33). There is still significant asymmetry (see Table 3.29) in terms of children's expectations exceeding parents' ($\chi_1^2 = 4.57$, $p < .05$) but the asymmetry is less marked. Parents are more optimistic about arithmetic than about reading; children are slightly less optimistic.

Conduct

The expectations for conduct are heavily skewed toward the high end for both parents and children. Eight parents look for a relatively poor showing (a "C") in conduct but none expects the worst mark. For 6 of these 8 worried parents, the child expects an outstanding performance! There are 10 children worried about conduct, but they are different children from the eight whom the parents are worried about. Again there is no significant matching between expectations (34% observed, 32% expected by chance; with $r = .03$) and significantly less optimism of parents compared to children ($\chi_1^2 = 8.32$, $p < .01$).

To sum up: Parents and children do not exhibit above chance levels of agreement in terms of expectations. Children consistently are more optimistic than parents on the average. Parents' expectations are very much alike across the areas of reading, arithmetic, and conduct, but children are noticeably less confident in arithmetic than in reading or conduct.

First Report Card Marks Compared to Parents' Expectations

Reading

Parents are better forecasters than children of marks children will receive in reading. In 50 of 83 instances (60%) there is agreement which is significantly above chance ($z = 3.37$, $p < .01$) (see Table 3.41). This represents an excess of 15% over what would be expected by chance. The most interesting discrepances occur for those cases where the parent expects a 2--in 60% of these the child actually gets a 2, but in almost 30% (17 cases) the child comes home with a 1. In the other 10% of these cases the child comes home with a 3. Parents as a group are good at forecasting the numbers of poor marks given--about 12% C's and no D's.

Arithmetic

The first arithmetic mark likewise is much closer to parents' expectations than to children's--again a high percentage (61%) of parents have their guesses exactly confirmed, but the level of significance ($z = 2.18$, $p < .05$) and the excess matching over chance (10%) are both less than was true for reading. In this area there is no marked imbalance between what is predicted by parents and how children perform in the direction of children outstripping parents' hopes. The same number of children (16) do better than parents expect as do worse (see Table 3.42).

Table 3.41

		Reading Mark, Time 1				
		1	2	3	Total	Percent
Parent's Reading Expectation, Time 1	1	11	3		14	16.9
	2	17	36	7	60	72.3
	3	1	4	3	8	9.6
	4		1		1	1.2
	Total	29	44	10	83	
	Percent	34.9	53.0	12.0		100.0

Table 3.42

		Arithmetic Mark, Time 1				
		1	2	3	Total	Percent
Parent's Arithmetic Expectation, Time 1	1	10	12	1	23	27.7
	2	7	41	3	51	61.4
	3	1	8		9	10.8
	Total	18	61	4	83	
	Percent	21.7	73.5	4.8		100.0

Table 3.43

		Conduct Mark, Time 1					
		1	2	3	4	Total	Percent
Parent's Conduct Expectation, Time 1	1	12	3	1		16	20.5
	2	21	28	5		54	69.2
	3		2	5	1	8	10.3
	Total	33	33	11	1	78	
	Percent	42.3	42.3	14.1	1.3		100.0

Conduct

In conduct parents correctly anticipate their child's mark in 58% of the cases while only 39% would be expected to do so by chance ($z = 3.60$, $p < .01$). In addition, when parents' guesses disagree there is a significant tendency for them to underestimate their child's performance ($\chi^2_1 = 4.36$, $p < .05$).

To sum up: In all three marking areas about 60% of parents can forecast their child's mark exactly. In each area this degree of accuracy is significant in statistical terms but perhaps more important, the accuracy is impressive in absolute terms. In both reading and conduct parents' inaccuracy is apt to lie in under-estimating. Parents as a group seem aware of the marking distributions actually used by the first-grade teachers.

The Consistency of Parents' Expectations From First to Second Grade

In second grade 71% of the same parents who had given their expectations in grade one were asked again for expectations (their child now being in grade two). About half the parents (28 out of 59) shifted their expectations for reading from grade 1 to grade 2. Parents show a slight, but not a significantly greater than chance, likelihood of holding the same reading expectation at the two times ($z = 1.63$, N.S.). As noted earlier, upon first inquiry parents tended to "play it safe" by expecting a 2. Of the parents who had earlier expected a 2 (73%), now a little over half still expected a 2, but about a third looked for a 1 in second grade. At second grade a substantial number of parents then have raised their expectations although the total number of parents expecting 1's or 2's is approximately the same. All of those who looked for poor marks in first grade are looking for A's and B's in second grade, but about the same percentage of parents as before (10%) look for C's.

In arithmetic the marginal distributions for parents' first-grade and second-grade expectancies are almost the same, and there is significant consistency (63%) between parents' guesses from one year to the next ($z = 3.30$, $p < .01$). Ten have shaved their hopes by one grade interval, while nine have increased their hopes the same amount. The overall patterns are very similar from one year to the next.

Conduct looks similar to arithmetic including the high consistency (67%) from one year to the next ($z = 3.68$, $p < .01$). Twenty-five per cent of the parents who previously expected a 2 come to expect a 1 in grade two.

In sum, parents display a tendency to hold the same expectation from the first grade to the second (significantly so for arithmetic and conduct). No drastic marginal changes occur but slight upward movements in expectations are present for those previously expecting 2's in reading and conduct.

Table 3.44

		Parent's Reading Expectation, Time 3					
		1	2	3	4	Total	Percent
Parent's Reading Expectation, Time 1	1	8	2			10	16.9
	2	13	23	6	1	43	72.9
	3	2	4			6	10.2
	Total	23	29	6	1	59	
	Percent	39.0	49.2	10.2	1.7		100.0

Are Parents' Expectations for Second Grade Affected by Earlier Marks?

Reading

Parents' expectations in second grade match to a significant degree ($z = 3.74$, $p < .01$) the performances their children have earlier displayed. In fact 61% of the parents' expectations match earlier marks, compared to the 38% consistencies expected by chance. Those whose expectations differ from the previous mark are about equally divided in direction--16% look for improvement and 23% look for a decline.

Arithmetic

The picture for arithmetic is the same as for reading in that there is considerable concordance between parents' expectations for second grade marks and the marks actually received at the end of first grade ($z = 2.09$, $p < .05$); 54% of parents believe their child will perform at the same level in arithmetic as earlier (41% would be expected to do so by chance). There is an interesting asymmetry though, in that of those parents who forecast a different mark (46%) 11% look for a better mark but 35% look for a poorer mark! This asymmetry is moderately significant ($\chi^2_1 = 6.03$, $p < .05$). Parents thus manifest considerably less confidence in their children's arithmetic performance than in reading performance.

Conduct

Middle class parents are realists when it comes to conduct, or at least they seem to use the same marginals teachers use in evaluating conduct. There is a highly significant amount of matching ($z = 4.56$, $p < .01$)--66% of the parents expect the same conduct mark in second grade as was obtained in first (compared to 39% matches expected by chance). Those who expect a different grade split about equally in each direction.

In sum, arithmetic is the only area that displays asymmetry--parents tending to expect lower marks than their children received at the end of grade one. All three areas display significant matching between the parents' second grade expectation and children's marks at the end of grade one. This matching may be thought of as either (a) the parents adopting of the earlier marks as their later expectation or (b) as a spurious relationship with parents' first grade expectation as the independent variable. (Previously it was shown that parents' grade one expectations matched their grade two expectations. If their first expectations also match the children's marks as the end of the first year (as they do for arithmetic and conduct in the combined analysis page 5-49) the observed matching could partially be a function of the fact that both variables now being considered matched with the parents' time 1 expectations.) The matching between parents' expectations at times 1 and 3 is not that strong, however (actually not significant in the case of reading), so

this explanation alone is probably insufficient to produce the observed matching there--especially in reading. It remains, however, a source that possibly contributes substantially to the observed matching between parents' second year expectations and arithmetic and conduct marks at the end of grade one.

The reader might also note that the causal efficacy of a discrepance between year end marks on changes in parents' expectations has not yet been examined. This awaits further analysis.

Sex Differences

Sex and Marks

First Marks

The major difference in first report cards by sex is that, of 10 children receiving 3's in reading, 9 are boys. (No 4's are given.) There is a fairly even division by sex in terms of arithmetic marks. In conduct 63% of the girls, compared to 24% of the boys received the highest mark (1). Only 7% of the girls receive a 3 in conduct, and none receives a 4. For boys, 22% receive a 3 and one boys receives a 4. Boys clearly receive lower first conduct marks.

Marks at End of First Grade

There is more skewing in opposite directions by sex at the time the second mark in reading is given. Whereas at midyear 1's were assigned about equally to boys and girls, at the end of the year about 61% of the girls, compared to 36% of the boys, receive the highest mark. There is as well some clumping of boys at the lower end of the distribution--more than 26% of the boys get the lowest two marks whereas only one girl receives a "3", the lowest mark assigned to a girl. The association between marks in reading and sex is moderately significant ($\chi^2_3 = 9.07$, $p < .05$) (see Table 3.45).

There is some slight superiority of girls in terms of arithmetic marks--all get 1's or 2's, whereas about 10% of the boys get 3's, but sex differences are small and non-significant.

In conduct boys again receive noticeably lower grades than girls. Whereas over 50% of girls receive a "1", less than 20% of boys receive a "1". In addition, only two girls get less than a "2" whereas almost one-third of the boys fall below a "2". Using only the categories "1", "2", and "3" for a comparison by sex, one finds a highly significant association ($\chi^2_2 = 15.39$) between sex and marks in conduct. The awarding of better conduct marks to girls continues a trend begun when first marks are given. The polarization by sex has increased over the first year. On first marking 16% of children received a "3" or less of whom about 4% are girls. On second marking about 20% receive a "3" or less of whom 2% are girls.

Table 3.45

		Reading Mark, Time 2					
		1	2	3	4	Total	Percent
Sex	Male	17	18	11	1	47	55.3
	Female	23	14	1		38	44.7
	Total	40	32	12	1	85	
	Percent	47.1	37.6	14.1	1.2		100.0

Table 3.46

		Arithmetic Mark, Time 2				
		1	2	3	Total	Percent
Sex	Male	17	25	5	47	55.3
	Female	17	21		38	44.7
	Total	34	46	5	85	
	Percent	40.0	54.1	5.9		100.0

Table 3.47

		Conduct Mark, Time 2					
		1	2	3	4	Total	Percent
Sex	Male	9	23	12	3	47	55.3
	Female	21	15	1	1	38	44.7
	Total	30	38	13	4	85	
	Percent	35.3	44.7	15.3	4.7		100.0

Marks at Midyear, Grade 2

By midyear second grade 22% of boys get a "3" or "4" in reading, approximately duplicating the percentage seen at the end of first grade. The performance of girls is not as good at midyear second grade as it was at the end of first grade however, because 16% get 3's at this time (only 3% got 3's at year end). The majority of girls get the highest mark, nonetheless, and clearly outshine boys in terms of superior marks.

Arithmetic marks also show some interesting shifts between first and second grade. At first the modal mark for both sexes is a "2" and little use is made of the "3" category (4% boys, 5% girls). Then at the end of grade one, although the modal mark for both sexes continues to be a "2", substantial numbers of both boys and girls get 1's (almost half the girls, and well over a third of boys). Eleven per cent of the boys get 3's while no girls do. At the middle of second grade, there is a sharp drop. Only 15% of boys and 18% of girls get 1's and higher percentages of both sexes get 3's (20% of boys and 24% of girls). There is thus considerable difference between first and second grades in how teachers assign arithmetic marks. One might think of a "honeymoon" year, followed by a thumping return to reality. In the second grade the picture for conduct remains about the same as it was in first grade, with perhaps a little less leniency toward girls.

Marks at the End of Second Grade

At the end of grade two the sex difference in reading marks has diminished and there are relatively few marks given to either sex below a 2 (only 7%). Girls still outperform boys in terms of A-marks (66% of girls get A's compared to 41% of boys).

In arithmetic boys are more variable in performance than girls. Over 28% of boys get A's, compared to 21% of girls. But 17% of boys get 3's and 4's compared to 11% of girls. The average mark by sex is 1.91 for boys compared to 1.89 for girls but boys are considerably more variable.

In conduct there is a strong trend for girls to get better marks. Over 60% of girls get A's and only 8% get C's (no D's). Boys, on the other hand, have a relatively large percentage getting C or D (20%) and only 28% receiving A's.

To sum up: Marks in reading show girls expending greater effort over the first two grades and receiving a disproportionate share of A's, starting from an approximately equal assignment of A's on the very first report card. Arithmetic marks show very little difference in average level by sex at any time, with the exception that boys' marks at times display more variability. In conduct girls outperform boys at every marking period, a finding which is not surprising.

Sex and Children's Expectations

First Grade

There are no strong sex differences in expectations for reading before the first marks are received. Girls are a little more hopeful for A's rather than B's. There is an interesting pattern in initial expectations for arithmetic, perhaps reflecting sex role socialization. About 21% of boys expect less than a B, compared to 24% of girls but whereas no boy expects the lowest grade, 7% of girls do. Thus a slightly higher percentage of girls expect a low grade and a few expect the worst. In conduct the expectations of the sexes are the same, and rather trite.

At the end of the first grade the expectations of both sexes are highly polarized at the high end for reading. The changes in arithmetic expectations are toward highly similar distributions for the sexes. In conduct 80% of girls expect the highest grade compared to something less than half of the boys. This expectation is veridical given the high percentage of A's in conduct assigned to girls.

Second Grade

In second grade high expectations continue to predominate for both sexes in reading and arithmetic, but there has been a noticeable moderation in reading expectations, the subject where children by now have the most feedback. At the end of first grade 58% of the boys and 67% of the girls expected an A. In the middle of second grade 48% of boys and 50% of girls expect an A in reading. In arithmetic children's expectations at the end of grade one and in the middle of grade two are highly similar and not much different by sex. The percentage of boys expecting low conduct marks increases from the end of first to the middle of second grade, form 17% to 26%, but the two sexes remain quite similar in terms of expectations.

By the end of second grade for the first time less than 50% of the boys expect an A in reading. Although a majority of the girls (55%) still look for an A in reading, this is a smaller majority than heretofore. The sexes still do not differ significantly in reading expectations. Likewise there has been considerable moderation in boys' expectations for arithmetic even though the modal category (47%) is A. For girls there also is a noticeable moderation, the modal category (50%) now being a B. The sexes again appear quite similar. In conduct there are slightly fewer low expectations for both sexes at the end of second grade. Once again the sexes remain similar.

In arithmetic children's expectations at successive time points continuously have some low cases (at the end of second grade 11% of the boys and 5% of girls still look for the lowest two grades). This same trend can be seen in every earlier table for arithmetic expectations--some children are pessimistic. There seems to be much less pessimism for reading--only 3% of the girls and 7% of boys look for a 3 or less in reading at the end of second grade. At the end of first grade 2% of girls and 4% of boys looked for less than a "B" in reading, whereas 13% of boys and 12% of girls looked for low marks in arithmetic at that time.

With small numbers of cases conclusions are not firm but both teachers and students seem to define performance very differently in reading and arithmetic. The different, and more rigorous, marking standards for arithmetic have earlier been noted, as has the negative association between expectations in the two areas.

There is of course considerable tolerance in everyday life for persons who are "poor" at arithmetic but relatively little tolerance for failures in reading. Some persons, in fact, announce with resignation, or even pride, that they "cannot add 1 and 1" but one never hears acknowledgment of difficulty in reading. It may be that children in this study reflect the mores of the larger society in their scaling down of hopes for achievement in arithmetic and in their openness in acknowledging a low level of aspiration. They do not seem to have the same attitude toward reading achievement.

As a general summary statement the most consistent observation is that the sexes do not differ in expectations over the two year period observed (conduct at time 2 is the only exception).

Sex and Parents' Expectations

Parents have very slightly higher expectations for girls' initial reading marks than for boys--about 16% of boys are expected to receive less than a "B" compared to 6% of girls. Since fewer boys than girls expect higher marks the degree of congruence between parents and children of both sexes is reasonably similar (36% and 25% matching for boys and girls respectively).

Surprisingly, there is only a slight margin in parents' expectations for high arithmetic marks favoring boys (30% vs. 25% expect A's for boys and girls respectively). There is less discrepance, however, between the sexes in a comparison of parents' and children's expectations for arithmetic than was the case for reading. About 28% of girls' parents have expectations that match their daughters' expectations for arithmetic compared to 32% of parents having expectations that match their sons'.

Parents' expectations for conduct are different for the two sexes--girls are expected to behave better than boys. Whereas 32% of girls' parents look for the highest grade in conduct, only 10% of boys' parents expect this. Rather surprisingly there is about the same low degree of congruence between parents' and children's conduct expectations by sex, with 37% matches for boys and 35% matches for girls.

Parents seem to display for each sex no more than chance matching of expectations--as one might have expected given that parents' expectations do not match children's expectations in general (see the discussion of parents' expectations).

In sum, parents do not seem to hold different expectations for boys and girls except possibly in conduct.

Sex, Marks and Parents Expectations

In terms of actual marks received in reading, all but 13% of boys do as well or better than their parents expect in reading. The figure for girls is 11%. What parents expect and marks actually received by the sexes are comparable.

In arithmetic just under 25% of the boys do worse than parents expected, while this is true for only 13% of girls. The distributions of marks as awarded are similar for boys and for girls but as mentioned above, parents do seem to have slightly higher hopes for boys than for girls in arithmetic. Overall, though, it seems that the sexes are comparable in the relations displayed between arithmetic marks and parents' expectations.

Parents expect boys to be less well-behaved than girls (as mentioned, 33% of girls and 10% of boys are expected to get A's in conduct) but teachers award many more high marks to girls in conduct than to boys (61% vs. 26% get A's). The congruence between parents' expectations and marks received in conduct turns out to be roughly equivalent for the two sexes because the marks awarded are similarly imbalanced by sex.

In sum the relationship between parents' expectations and marks in all three areas are reasonably comparable between the sexes despite sex differences in marks and moderate differences in expectations.

Sex, Marks and Children's Expectations

Girls are more optimistic than boys about the first mark in reading (68% vs. 52% expect A's) and teachers give a few more A's--and noticeably more C's--to boys in reading on the first report card. Over 18% of boys get C's compared to a single girl (4%) getting a C. Almost half the girls (45%) get a lower mark in reading than they expect, and a large percentage of boys (39%) also get a mark lower than they expect. What is striking, though, is that about 25% of the boys who expect an "A" actually get a "C"--this does not happen to girls as frequently (4%).

In arithmetic similar findings prevail in terms of marks compared to expectations--half the boys get a lower mark in arithmetic than they expected and 39% of girls do. There is no marked increase in low marks for boys as occurred for reading however.

In conduct boys and girls are very optimistic, but the girls' optimism is more fully supported by later events than the boys'. Teachers give only 8% of girls a conduct mark below a "B" compared to 23% of boys. With this almost a third of the girls get a higher mark in conduct than they expected whereas only about 18% of boys get a higher mark than expected.

The chance levels of matching between first marks and expectations observed in the whole cohort are evident in all three areas for both boys and girls.

Relations with I.Q.

While a significant relationship between I.Q. (PMA) and reading is present at the middle of first grade ($r = .224$, $p < .05$) it is not until the middle of second grade that significant relationships between I.Q. and marks in reading and arithmetic again reappear ($r = .273$, $p < .01$ and $.374$, $p < .01$, respectively). These relations fall off (to non-significance in the case of reading) at the end of second grade, however, whether the early (PMA) or later (SFTAA from STMM) I.Q. measure is used (PMA: $r = .145$, N.S., $r = .196$, $p < .05$, and SFTAA: $r = .131$, N.S., $r = .201$, $p < .05$ for reading and arithmetic respectively). There is a sizeable correlation (.655) between I.Q. measured in first and second grades using the two different tests.

Because the marks are not a continuously distributed variable, and because the large majority of cases fall in only the two uppermost mark categories, the relationships between I.Q. and marks may be attenuated to a considerable extent.

If quartiles are formed on the I.Q. distribution (the necessary cut points being 124.5, 116.5, and 109.5) there is little difference in the distributions of first reading marks within the two middle categories and there are moderate differences within the two extreme categories. Of the 19 cases in the upper I.Q. quartile 58% get 1's and 42% get a lower mark. Of the 20 cases in the lowest quartile 20% get 1's and 80% get a lower mark. To give some notion of the lack of relation between marks and I.Q., at the end of first grade those children receiving the highest mark ("1") in reading, come from the different quartiles (from highest to lowest) in the following proportions: .24, .14, .38, .24. The picture is very similar in arithmetic--.21, .12, .43, .24.

By the middle of second grade, there is a little more consistency in terms of I.Q. quartile and mark received. This is most noticeable in arithmetic where of those children receiving a "1" in arithmetic, the quartile percentages are as follows: 46%, 15%, 15%, 23%. Of those children receiving a "3" in arithmetic, the quartile percentages are: 0%, 18%, 24%, 59%. At the end of second grade the quartile frequencies are less "consistent" than they were in the middle of the second grade year--reflecting the decline in correlation.

The agreement between I.Q. quartiles for grades 1(PMA) and 2 (SFTAA) is best at the high extreme (see Table 3.52). About 86% of those classified in the highest quartile in first grade are classified the same way again, and the remaining 14% are reclassified into the second quartile. Of those classified in the lowest quartile in grade 1, a little less than half (46%) receive the same quartile rating in grade 2, but 8% have jumped to the top quartile. But of those classified in the third quartile in grade 1--the quartile that shows the greatest discrepance upon retesting in grade 2--over a third have jumped to the top quartile by grade 2 and 13% have slipped to the lowest quartile at the time of second testing.

These distributions illustrate a fact well-known but little appreciated: with young children high I.Q. scores are more apt to be accurate than low ones. After all, there is almost no way that a child

Table 3.48

		Reading Mark, Time 1				
		1	2	3	Total	Percent
Categorized I.Q. (PMA) in Grade 1	1st = High	11	6	2	19	23.2
	2nd	5	11	3	19	23.2
	3rd	10	14		24	29.3
	4th = Low	4	11	5	20	24.4
	Total	30	42	10	82	
	Percent	36.6	51.2	12.2		100.0

Table 3.49

		Reading Mark, Time 3					
		1	2	3	4	Total	Percent
Categorized I.Q. (PMA) in Grade 1	1st = High	8	5	1		14	17.7
	2nd	6	9	4		19	24.1
	3rd	9	10	3	1	23	29.1
	4th = Low	5	12	3	3	23	29.1
	Total	28	36	11	4	79	
	Percent	35.4	45.6	13.9	5.1		100.0

Table 3.50

		Arithmetic Mark, Time 3				
		1	2	3	Total	Percent
Categorized I.Q. (PMA) in Grade 1	1st = High	6	8		14	17.5
	2nd	2	14	3	19	23.7
	3rd	2	17	4	23	28.7
	4th = Low	3	11	10	24	30.0
	Total	13	50	17	80	
	Percent	16.2	62.5	21.2		100.0

Table 3.51

		Arithmetic Mark, Time 4					
		1	2	3	4	Total	Percent
Categorized I.Q. (SFTAA) in Grade 2*	1st = High	11	13	3		27	26.5
	2nd	3	22	4		29	28.4
	3rd	6	15	1		22	21.6
	4th = Low	3	14	6	1	24	23.5
	Total	23	64	14	1	102	
	Percent	22.5	62.7	13.7	1.0		100.0

* (cut points at 111.5, 103.5, and 94.5)

Table 3.52

		Categorized I.Q. (SFTAA) in Grade 2					
		High = 1st	2nd	3rd	Low = 4th	Total	Percent
Categorized I.Q. (PMA) in Grade 1	1st = High	12	2			14	17.5
	2nd	5	10	3	1	19	23.7
	3rd	8	7	5	3	23	28.7
	4th = Low	2	4	7	11	24	30.0
	Total	27	23	15	15	80	
	Percent	33.7	28.7	18.7	18.7		100.0

251

can score much higher than his ability, but there are innumerable ways he can score lower (misunderstanding directions, inattention, anxiety, poor vision, etc.).

While the correlation between I.Q. measures at the two times is sizeable (.655) it is not as high as would be preferred for a retest. With this low a reliability, true score I.Q.-mark correlations may have some difficulty displaying themselves.

It should be noted in the interpretations of the above results that the marking policy of the middle-class school (marking students in terms of their own ability) has the effect that teachers are attempting to partial out I.Q., or an I.Q. like component, from the marks they assign. To the extent they are successful I.Q.-mark correlations will be diminished. The presence of some significant I.Q.-mark correlations indicates teachers are not totally successful at partialling out ability--at least "ability" as measured by the I.Q. tests the teachers had access to--but the extent to which they are successful remains largely unknown.

Relation Between Self-Esteem Test and Expectations

The self-esteem test developed by Dickstein (1972) has already been described in Chapter 2. The reader will recall that it has two forms, one for boys and one for girls. The correlations that follow are generally based on from 40 to 45 boys and 31 to 37 girls.

Boys

No correlations of any of the three self-esteem factors and marks received at midyear first grade are significant at the 5% level for boys, although all are positive (9 coefficients). Factor II (athletic image) does correlate in excess of .20 with all three marks.

Of the nine correlations between boys' midyear (grade one) expectations in the three areas and the self-esteem factors, the only correlation that attains significance ($p < .05$) is that between conduct expectation and factor III (student image). For boys, expecting a good conduct mark correlates positively with holding a good student image (.269)--which is not too surprising. It is more surprising that the student image factor does not correlate significantly with academic expectations (.213 and .072 for reading and arithmetic respectively).

At the end of first grade there is a highly significant correlation between factor III (student image) and boy's mark in reading ($r = .404$, $p < .01$), and a moderately significant correlation with boys' mark in arithmetic (.267, $p < .05$). Student image also correlates significantly with arithmetic expectations at year end (.254, $p < .05$) but not with reading expectation (.182, N.S.). Factor I (good citizen image) correlates with boys' expected mark in conduct ($r = .341$, $p < .05$) at year end.

There is no significant relation between I.Q. and the three self-esteem factors though all are positive and factor I (good citizen image) comes very close to significance ($r = .257$).

There is a significant correlation between the scores on factor I as measured in first grade and factor I as measured in second grade--a 12 to 15 month time span ($r = .357$, $p < .05$). The over-time correlation for factor II is much more substantial--.520--and also highly significant ($p < .01$) as is factor III (.437).

Factor III, the student image as measured in grade 2, again shows a significant correlation with the marks received by boys in reading ($r = .285$, $p < .05$, $n = 53$) and the marks received in arithmetic ($r = .297$, $p < .05$). In addition, factor I (good citizen image) shows fairly sizeable positive correlations with all three marks at the end of grade 2, with those for reading and conduct being significant (.276 and .310, respectively).

Girls

The same kind of relationships between self-esteem factors and school variables were tested for girls.

At first marking (midyear, first grade) there were no discernible relationships between any of the self-esteem factors and girls' marks in the three criterion subjects. At the end of first grade, some sizeable relationships began to surface. Factor II (athletic image) then correlates with both reading and arithmetic marks ($r = -.335$, $p < .05$ and $-.404$, $p < .05$) and factor III (social image) correlates with arithmetic ($r = -.403$, $p < .05$) but these are all in the opposite direction compared to boys. For girls, low self-exteem on the factors correlates with high marks! The non-significant correlations at the end of grade one are mixed as to their sign. There is only one significant correlation between girls' expectations and self-esteem factors at the middle of first grade, that of athletic image (factor II) and reading expectation ($-.308$, $p < .05$). It is in the direction of high self-esteem being associated with a low expectation. At the end of first grade factor II (athletic image) continues to correlate in this fashion with reading expectations ($r = -.325$, $p < .05$). At both the middle and the end of grade one the non-significant correlations are mixed as to sign.

As far as I.Q. is concerned, there is no significant association between it and any of the self-esteem factors early in first grade.

There is no association between scores on factor I (good citizen image) as measured early in first grade and as measured at the end of second grade. There is a similar lack of correlation for factor III (social image). But there is a significant ($r = .381$, $p < .05$) association for factor II (athletic image). At the end of second grade, there are still no significant correlations between self-esteem (as measured in grade 2) and any marks for girls.

In summary boys' and girls' marks and expectations display quite different relations to the various self-esteem factors--including changes in both sign and magnitude of correlations.

CHAPTER 4
MIDDLE-CLASS SCHOOL, COHORT 2
(STARTING FIRST GRADE IN SEPT. 1972)

A second cohort at the middle-class school was followed starting in September, 1972, consisting of all children beginning first grade in that school at that time. Complete data for these children cover the school year 1972-73, although some prior data in terms of kindergarten teachers' forecasts were obtained. These children are presently being followed in second grade and results will be available in future reports. They have demographic characteristics like those for Cohort 1 in the same school (see Table 3.1).

This chapter is short because most of the data pertaining to Cohort 2 are given in Chapter 5 (combined with data for Cohort 1). In this chapter only a few points will be covered where there seem to be differences between cohorts or where the details of cohort-to-cohort comparisons are illuminating.

Tabular Summary

In reading parents and children expect virtually the same performance (1.80 vs. 1.79 respectively). In arithmetic parents are less optimistic than children (2.03 vs. 1.94), and both are somewhat more guarded about arithmetic as compared to reading. For conduct parents are more optimistic than children (1.97 vs. 2.12). As with Cohort 1, the variability of children's forecasts exceeds parents'. Compared with Cohort 1, parents look very much the same (their expectations ranging from 1.80 to 2.03 in Cohort 2 vs. 1.83 to 1.96 in Cohort 1). Children of Cohort 2 are less optimistic than children of Cohort 1 (ranging from 1.79 to 2.12 vs. 1.63 to 1.68). (See Table 5.1 for a complete summary).

Discrepances Between Initial Expectations and First Report Card

Reading

The teachers, first of all, give almost all A's and B's at the first marking opportunity; 56% of the children get B's and 34% get A's. The children, by contrast, expect many more A's (over half of them expect A's) and many more expect a relatively low mark than actually are awarded low

Table 4.1

		Reading Mark, Time 1				
		1	2	3	Total	Percent
Reading, Expectation, Time 1	1	15	28	3	46	51.1
	2	11	9	2	22	24.4
	3	3	11	3	17	18.9
	4	2	2	1	5	5.6
	Total	31	50	9	90	
	Percent	34.4	55.6	10.0		100.0

Table 4.2

		Arithmetic Mark, Time 1				
		1	2	3	Total	Percent
Arithmetic Expectation, Time 1	1	4	18	3	25	27.8
	2	12	28	9	49	54.4
	3	1	8	2	11	12.2
	4		2	3	5	5.6
	Total	17	56	17	90	
	Percent	18.9	62.2	18.9		100.0

Table 4.3

		Conduct Mark, Time 1					
		1	2	3	4	Total	Percent
Conduct Expectation, Time 1	1	10	21	4		35	38.9
	2	4	8	6		18	20.0
	3	6	16	5		27	30.0
	4	2	6	1	1	10	11.1
	Total	22	51	16	1	90	
	Percent	24.4	56.7	17.8	1.1		100.0

marks. About 25% expect a C or D while only 10% get C's, and none lower. The greatest difference between this Cohort and Cohort 1 lies here--for only 10% of the children in Cohort 1 expected a C or a D in reading initially. The predicted numbers of matches between expectations and marks in reading, 29.77, is close to the number of matches observed, 27, so the extent of of matching between initial expectation and marks in reading is random (Table 4.1).

It is interesting to examine the 2 X 2 array in the upper left corner of Table 4.1 where more than two-thirds of the cases fall. Children's expectations for A's and B's are not consonant with marks received--of the large numbers (43) who expect A's, the majority receive B's. Of those expecting B's (20) the majority receive A's. This situation almost exactly duplicates what was seen in Cohort 1.

Arithmetic

Children in Cohort 2 are considerably less optimistic about arithmetic than children in Cohort 1. Although 82% look for A's and B's, only 28% look for an A. This compares with 89% looking for A's or B's in Cohort 1, with 61% looking for an A. Compared on expectations for reading children of Cohort 2 are much more guarded in arithmetic, while children of Cohort 1 were more optimistic about arithmetic. Also, there is a different division between A's and B's in the different substantive areas. In arithmetic more than 50% look for B's whereas in reading the modal (and majority) expectation was for an A.

Teachers do mark arithmetic somewhat more severely, because in Cohort 2 12% get C's whereas in Cohort 1 only 4% received C's.

The amount of matching between children's expectations and marks is within sampling error of the 37.29 matches predicted by chance (34 matches are observed). There is also a fairly even division between those who get a higher grade than expected and those who get a lower grade than expected.

Conduct

The picture here is not particularly interesting. The amount of matching (24) is well within chance expectancy (23.7), there is no trend for expectations to exceed marks or the reverse, but teachers give noticeably fewer A's in conduct to Cohort 2 (24%) than to Cohort 1 (43%). Also, interestingly, children in Cohort 2 have noticeably lower expectations for conduct (59% look for A's or B's compared to 87% in Cohort 1).

Kindergarten Teachers' Forecasts

Since teachers' expectations are often assumed to influence future performance, it is of interest to see how well kindergarten teachers' forecasts for performance in first grade are borne out. These data are

Table 4.4

		Kindergarten Teacher's Reading Expectation					
		1	2	3	4	Total	Percent
Reading Mark, Time 1	1	3	16	3		22	36.1
	2	10	12	9	4	35	57.4
	3	1	2	1		4	6.6
	Total	14	30	13	4	61	
	Percent	23.0	49.2	21.3	6.6		100.0

available only for Cohort 2. Forecasts were obtained by asking kindergarten teachers to guess how well children would do in first grade. The guesses were obtained at the end of kindergarten, just before school closed, after the teacher had spent 8 to 9 months in fairly close association with the children.

The forecasts are far off the mark. There is less congruence than would be expected by chance. Whereas 23.11 exact matches would be predicted, only 16 are noted ($z = 2.04$, $p < .05$). Also the kindergarten teachers are considerably less optimistic than the children's performance warrants, and also less optimistic than parents are. They look for about 72% to get A's and B's, and over 93% of the children actually do so. Also 32% of parents look for A's whereas fewer (23%) of kindergarten teachers do. The kindergarten teachers' lack of optimism is shown in the departure from symmetry around the main diagonal ($X_1^2 = 7.20$, $p < .01$) brought about because more often than not, when expectations and marks received do not agree, it is the teacher's expectation which is lower. The kindergarten teachers seem to be less aware of the grading distributions used by the first grade teachers than parents are, for parents' marginals looked very much like the actual distribution of marks awarded in reading whereas kindergarten teachers expected 28% to get C's or D's compared to 7% actually receiving these low marks.

What is the Correspondence Between Marks from Midyear to Year End?

Reading

In reading the teachers' marginals are fairly stable from one marking period to the next for Cohort 2. If anything a little more leniency appears as the year goes on. Over 91% get A's and B's on second marking. In 62% of the cases the mark awarded stays the same--a degree of over-matching that is highly significant ($p < .01$). There are about equal numbers of children whose marks decline as there are those whose marks improve. In Cohort 1, on the other hand, teachers are noticeably more generous with A's in reading at the end of the year compared to the midyear (47% vs. 35%). This leads to a smaller percentage of matching marks (55%) on the two occasions for Cohort 1 than is the case for Cohort 2.

Arithmetic

In arithmetic the picture is more complex. There is a high degree of matching of arithmetic marks over the year (67%). This is significantly ($p < .01$) above the number of matches expected by chance (42%). Teachers tend to assign more A's at year end (36% vs. 18%) and fewer C's or D's (11% vs. 19%). In Cohort 1, more A's were assigned at year end (41% vs. 22%) and there were the same number of C's (5%) both times.

Table 4.5

		Reading Mark, Time 2				
		1	2	3	Total	Percent
Reading Mark, Time 1	1	21	10		31	33.3
	2	9	36	7	52	55.9
	3		9	1	10	10.8
	Total	30	55	8	93	
	Percent	32.3	59.1	8.6		100.0

Table 4.6

		Arithmetic Mark, Time 2					
		1	2	3	4	Total	Percent
Arithmetic Mark, Time 1	1	15	1	1		17	18.3
	2	16	40	1	1	58	62.4
	3	2	9	7		18	19.4
	Total	33	50	9	1	93	
	Percent	35.5	53.8	9.7	1.1		100.0

Table 4.7

		Conduct Mark, Time 2					
		1	2	3	4	Total	Percent
Conduct Mark, Time 1	1	20	2			22	23.7
	2	13	37	3	1	54	58.1
	3	1	10	5		16	17.2
	4			1		1	1.1
	Total	34	49	9	1	93	
	Percent	36.6	52.7	9.7	1.1		100.0

In Cohort 2 the degree of leniency for arithmetic marks as the year progresses looks about equal to that observed for reading--81% get A's and B's at first mark and 89% do so at second mark. There is however, a very different pattern around the main diagonal. Whereas slightly over 4% of children's marks in arithmetic go down, 31% go up. There are, in addition two children whose marks fall by 2 notches and 2 whose marks rise by two notches. Because the teachers' grading distributions are initially "harder" in arithmetic, they can award twice as many A's at the second mark as they do at the first. In reading, by contrast, the same proportion of A's is awarded on both occasions.

Conduct

The pattern over time for marks in conduct closely resembles that for arithmetic. Conduct also displays a significant over-matching of marks at midyear and year end (see Table 4.7).

It is rather surprising that teachers' marks differ as much as they do from Cohort 1 to Cohort 2.

Parents' Expectations

At the beginning of this chapter, data on parents' expectations were reviewed (See also Table 5.1) and it was pointed out that parents of children in Cohort 2 have about the same expectations as parents of children in Cohort 1.

Parents' Expectations Compared to Children's Expectations

Reading

The degree of congruence between the expectations of parents and children is only what would be predicted in terms of chance. The expected number of matches is 23, with a standard deviation of 3.5. In 19 out of 73 cases (26%) for reading there is perfect agreement (in Cohort 1 the comparable figure is 31%), well within chance expectation. Unlike Cohort 1 where children were noticeably more optimistic than parents, in Cohort 2 there are equal numbers of disagreements in both directions. In 27 instances parents are less optimistic than children and in 27 instances parents were more optimistic. The patterns of disagreement, however, in the optimistic and pessimistic directions are very different.

In almost all of the instances where parents are less optimistic than children, the child looks for a 1 while the parent looks for a 2 (22 of 27 instances). The marginals for parents and children are not alike. Over 50% of children look for the highest grade while only about

Table 4.8

		Parent's Expectation, Reading Time 1					
		1	2	3	4	Total	Percent
Child's Expectation, Reading, Time 1	1	11	22	4		37	50.7
	2	9	7			16	21.9
	3	4	10	1	1	16	21.9
	4	1	1	2		4	5.5
	Total	25	40	7	1	73	
	Percent	34.2	54.8	9.6	1.4		100.0

Table 4.9

		Parent's Expectation, Arithmetic, Time 1					
		1	2	3	4	Total	Percent
Child's Expectation, Arithmetic, Time 1	1	8	12	3		23	31.5
	2	8	22	9		39	53.4
	3		5	2		7	9.6
	4		2		2	4	5.5
	Total	16	41	14	2	73	
	Percent	21.9	56.2	19.2	2.7		100.0

Table 4.10

		Parent's Expectation, Conduct, Time 1					
		1	2	3	4	Total	Percent
Child's Expectation, Conduct, Time 1	1	5	16	4	1	26	35.6
	2	6	8	2		16	21.9
	3	5	15	3	1	24	32.9
	4	1	4	2		7	9.6
	Total	17	43	11	2	73	
	Percent	23.3	58.9	15.1	2.7		100.0

one-third of parents do. Similarly a large percentage (55%) of parents have modestly optimistic expectations--a "2"--but 22% of children hold this expectation. Over 27% of children look for a 3 or a 4, compared to 11% of parents. Children's expectations are thus considerably more variable than parents'. Children of both Cohort 1 and Cohort 2 hold more extreme expectations than parents. In Cohort 2 the variability operates in both directions (some children expect very low grades) whereas in Cohort 1 there was over-optimism only.

Arithmetic

There is more congruence between parents' and children's expectations for arithmetic (47% exact agreement) than was true for reading, but again a matching test indicates this amount of agreement is not significant given the marginals. What appears to be greater congruence stems from the similar marginals of the two groups. Parents and children look for the same number of 2's (56% and 53% respectively) and the only major difference between them is an imbalance between 1's and 3's. About 31% of children expect 1's compared to 22% of parents. About twice as many parents look for 3's as children (19% vs. 10%). Somewhat more pessimism is registered by parents than by children (33% of parents are less optimistic about arithmetic than their child, compared to 21% being more optimistic) but the differential optimism is not extensive enough to be statistically significant ($\chi^2_1 = 2.08$, N.S.).

A difference between parents in Cohort 1 and those of Cohort 2 is that parents of Cohort 2 differ in their marginals for reading and arithmetic, expecting higher reading grades. Parents in Cohort 1 held about the same expectations in each area.

Conduct

Expectations for conduct are remarkable both because of children's pessimism--43% of them expect a "3" or a "4"--and because of what appears to be a lower degree of congruence between parents and children (22%) than was true for arithmetic. Actually given the two marginal distributions, the degree of congruence is well within the range of what would be predicted by chance. Again parents pick the "safe" alternative, a "2", and again children resort to the extremes of the distribution more often. As well as many expectations for 3's and 4's, about 36% of the children look for 1's. The asymmetry in the table points toward children being less optimistic than parents but this imbalance is not statistically significant ($\chi^2_1 = 1.12$, N.S.).

Discrepance Between Parents' Initial Expectation and First Report Card

Reading

There is remarkable overall concordance at the aggregate level between parents' marginal distribution and the distribution of teachers'

Table 4.11

		Reading Mark, Time 1				
		1	2	3	Total	Percent
Parent's Expectation, Reading, Time 1	1	8	14	1	23	32.4
	2	14	24	1	39	54.9
	3	1	4	3	8	11.3
	4		1		1	1.4
	Total	23	43	5	71	
	Percent	32.4	60.6	7.0		100.0

Table 4.12

		Arithmetic Mark, Time 1				
		1	2	3	Total	Percent
Parent's Expectation, Arithmetic, Time 1	1	4	11	1	16	22.5
	2	10	26	3	39	54.9
	3	2	6	6	14	19.7
	4			2	2	2.8
	Total	16	43	12	71	
	Percent	22.5	60.6	16.9		100.0

Table 4.13

		Conduct Mark, Time 1				
		1	2	3	Total	Percent
Parent's Expectation, Conduct, Time 1	1	5	10	1	16	22.5
	2	8	27	8	43	60.6
	3	2	5	3	10	14.1
	4		1	1	2	2.8
	Total	15	43	13	71	
	Percent	21.1	60.6	18.3		100.0

marks--the same number of 1's are awarded as parents expect, no 4's are given as parents expect, and the numbers of 2's and 3's awarded are close in the two distributions. In other words the pattern of marks awarded conforms with what parents overall expect. On an individual level, however, as particular parents and children are matched, there is not much agreement between expectations of parents and children's first reading mark. There is a little more, but not significantly more, concordance between parents expectations for reading marks and children's performance than would be expected by chance (35 cases (out of 71) of agreement--just about 50%--where the expected agreement is 31.63). Thus 23 parents look for A's and 23 children are awarded A's, but in only 8 instances do the awards and parents' hopes for A's coincide.

Arithmetic

There is concordance at the aggregate level for arithmetic marks also, the marginals for parents' expectations looking very much like the marginals for marks awarded. At the individual level, over 50% of the expectations are exactly matched by the marks awarded--as was the case in reading--but since the number expected under the matching model is 29.59 for arithmetic with a standard error of 3.51, the observed value of 36 only approaches significance ($z = 1.68$, $p < .10$). There is thus slightly more agreement between what parents expect and what happens in arithmetic.

Conduct

In conduct the concordance between parents' expectations and marks received is of the same variety as in reading or arithmetic and not significantly different from chance. The mark and parent expectation marginals are again very similar.

In summary, the children of Cohorts 1 and 2 hold somewhat different expectations initially and teachers' marks for Cohort 2 seem more severe, especially in reading. Parents' expectations are very similar from cohort to cohort, and this leads to a mild disagreement between expectations of parents and children in Cohort 1 but not in Cohort 2. Surprisingly, kindergarten teachers' forecasts are poor predictors of performance in first grade, significantly worse than chance level.

CHAPTER 5
RESULTS FOR TWO FIRST-GRADE COHORTS (1971-72, 1972-73), WHITE MIDDLE-CLASS SUBURBAN SCHOOL

This analysis permits a summing up of findings present in the two white middle-class suburban cohorts, and in some instances, an opportunity to look for effects that were too small or too complex to be clearly shown in the separate cohorts. As mentioned, we have assumed time-of-sampling to be negligible, that is, we assume differences between cohorts reflect differences in groups of children making up the cohorts rather than differences attributable to time-of-sampling--fall, 1971 vs. fall, 1972. If results for the two cohorts conflict, our conclusion will be that the findings are uninterpretable with respect to developmental change. Where findings agree, data have been combined so associations and trends can be more reliably measured, and we assume that such findings do represent developmental trends.

Tabular Summary

Means and standard deviations for the variables used in subsequent analyses and cross-tabulations are given in Table 5.1. The mean IQ for 183 children (those in the sample, 82% for whom IQ data were available from school records) is 113.5, with a standard deviation of 11.

A number of conclusions, borne out in both cohorts separately, are even clearer in the data for combined cohorts.

(1) Expectations were obtainable from 86% of parents. All were contacted at least twice. Parents have lower expectations than children for reading and arithmetic, but only the difference for reading approaches significance ($t = 1.95$ for the combined cohorts). Parents' expectation levels are uniform across subject areas when the two cohorts are combined.

(2) Children in Cohorts 1 and 2 show significant differences in their expectations, those in the second cohort being lower. For reading the difference (0.16) is small in absolute terms and not significant ($t = 1.17$). For arithmetic the difference is still fairly small (0.28, about a quarter of a grade point) but significant ($t = 2.26$, $p < .05$). For conduct the difference is largest, being 0.45, and this difference is highly significant ($p < .01$). The differences between cohorts are such as to make children's and parent's expectations more consistent in Cohort 2 than in Cohort 1.

(3) At the middle of grade one children have highest expectations in reading; next highest in arithmetic, and least in conduct. These differences between subject areas are consistent in both cohorts, but

Table 5.1

Means, Standard Deviations for Separate Cohorts and Sum of Cohorts During First-Grade Year
Middle-Class School

	Cohort 1			Cohort 2			Combined Cohorts		
	N	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.
Parents' Expectations--T1									
Reading	84	1.96	0.57	76	1.80	0.67	160	1.89	0.62
Arithmetic	84	1.83	0.60	76	2.04	0.74	160	1.93	0.67
Conduct	79	1.89	0.55	75	1.97	0.70	154	1.93	0.63
Child's Expectation --T1									
Reading	90	1.63	0.90	96	1.79	0.96	186	1.72	0.94
Arithmetic	90	1.66	0.90	96	1.94	0.79	186	1.80	0.86
Conduct	90	1.68	0.90	96	2.13	1.05	186	1.91	1.00
Child's Expectation--T2									
Reading	90	1.42	0.60	94	1.65	0.60	184	1.54	0.61
Arithmetic	90	1.64	0.81	94	1.75	0.75	184	1.70	0.78
Conduct	90	1.49	0.72	94	1.71	0.73	184	1.60	0.73
Child's Marks--T1									
Reading	85	1.77	0.65	93	1.77	0.63	178	1.77	0.64
Arithmetic	85	1.82	0.49	93	2.01	0.62	178	1.92	0.57
Conduct	85	1.75	0.75	93	1.96	0.67	178	1.86	0.72
Chils's Marks--T2									
Reading	86	1.69	0.76	94	1.77	0.59	180	1.73	0.68
Arithmetic	86	1.66	0.59	94	1.77	0.66	180	1.71	0.63
Conduct	86	1.88	0.83	94	1.76	0.67	180	1.82	0.75

Summary of Changes Over Year

	Cohort 1			Cohort 2			Combined Cohorts		
	N	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.
Mark Discrepance (T1 Mark minus T2 Mark)									
Reading	78	0.10	0.70	93	0.01	0.62	171	0.05	0.65
Arithmetic	78	0.19	0.56	93	0.25	0.64	171	0.22	0.60
Conduct	78	-0.04	0.63	93	0.20	0.60	171	0.09	0.63
Expectation Discrepance (T1 Expectation minus T2 Expectation)									
Reading	85	0.24	0.88	89	0.18	1.14	174	0.21	1.02
Arithmetic	85	0.00	1.14	89	0.23	0.84	174	0.12	1.00
Conduct	85	0.21	1.03	89	0.47	1.19	174	0.35	1.12
Mark-Expectation Discrepance T1 (T1 Mark minus T1 Expectations)									
Reading	82	0.18	1.07	90	-0.03	1.08	172	0.07	1.07
Arithmetic	82	0.18	0.97	90	0.04	0.91	172	0.11	0.94
Conduct	82	0.10	1.06	90	-0.18	1.20	172	-0.05	1.14
Mark-Expectation Discrepance T2 (T2 Mark minus T2 Expectation)									
Reading	84	0.27	0.87	91	0.11	0.71	175	0.19	0.79
Arithmetic	84	0.02	0.88	91	0.02	0.86	175	0.02	0.86
Conduct	84	0.37	0.92	91	0.03	0.90	175	0.19	0.92

only the reading-conduct difference reaches significance ($p < .05$) in the combined cohorts. Differences between areas are tiny in Cohort 1.

(4) Children's expectations in reading and conduct increase significantly over the first-grade year ($t = 2.60$, $p < .05$; $t = 3.60$, $p < .01$, respectively for the combined cohorts). Those for arithmetic increase also, but not to a statistically significant degree ($t = 1.74$, $p < .10$ for the combined cohort). The smaller increase for arithmetic expectations results in the children's arithmetic expectations being the lowest of their expectations at the end of first grade while conduct expectation was lowest in the middle of the first grade. Both cohorts individually show the same pattern of relationships, the only difference being that the second cohort shows a more pronounced increase in arithmetic expectations between midyear and year end than does the first cohort.

(5) Children's marks go up significantly in arithmetic over the year in both cohorts (Cohort 1, +0.16, $t = 3.20$, $p < .01$; Cohort 2, +0.24, $t = 3.75$, $p < .01$) as well as in the combined cohort (+0.21, $t = 4.93$, $p < .01$). Reading marks remain relatively stable over the year in both cohorts. Conduct moves down in the first cohort, up in the second and for the two combined is almost stable.

Before the First Report Card What Do Children Expect?

First, a word is needed about how marks are assigned. At the first marking period (around the end of November) first graders do not receive report cards. At the end of the first semester (around February 1) they receive a report. The card says: "This report is designed to measure the progress of your child in terms of his own maturity and ability. Comparison with other children or groups of children is avoided." Thus children are being judged in terms of what they are theoretically capable of doing--their performance, in other words, is assessed with IQ partialled out.

Discrepancies Between Initial Expectations and the First Report Card

As was true for each cohort taken separately there is not a significant degree of matching in the combined cohorts between children's expectations and the marks they actually receive on their first report card in reading, arithmetic, or conduct. (As can be seen in Table 5.1 the variability in expectations exceeds that in marks.) A high percentage of children expect A's (55% expect A's in reading, for example) and also a sizeable percentage (6%) expect failure. Actually no one fails. (In conduct 2 children are given failing grades, whereas 15 (9%) expect them.) As can be seen in Tables 5.2, 5.3, and 5.4, the outstanding characteristic in every area is the large number of children expecting 1's who get 2's, far outbalancing the number who get 1's but expect 2's. This kind of over-optimism is the principal source of disagreement between expectations and first marks.

Table 5.2

		Reading Mark, Time 1				
		1	2	3	Total	Percent
Reading Expectations, Time 1	1	30	55	10	95	55.2
	2	23	21	2	46	26.7
	3	3	13	4	20	11.6
	4	3	6	2	11	6.4
	Total	59	95	18	172	
	Percent	34.3	55.2	10.5		100.0

Table 5.3

		Arithmetic Mark, Time 1				
		1	2	3	Total	Percent
Arithmetic Expectations, Time 1	1	17	53	5	75	43.6
	2	14	41	9	64	37.2
	3	4	18	3	25	14.5
	4		5	3	8	4.7
	Total	35	117	20	172	
	Percent	20.3	68.0	11.6		100.0

Table 5.4

		Conduct Mark, Time 1					
		1	2	3	4	Total	Percent
Conduct Expectations, Time 1	1	33	35	11	1	80	46.5
	2	14	21	9		44	25.6
	3	8	19	6		33	19.2
	4	2	10	2	1	15	8.7
	Total	57	85	28	2	172	
	Percent	33.1	49.4	16.3	1.2		100.0

Discrepances Between Expectations and Report Cards at End of First Grade

By the end of the first grade children show some improvement in their ability to forecast marks accurately. For reading 53% of children ($z = 2.82$, $p < .01$) are correct, and for arithmetic 46% ($z = 1.65$, $p < .10$) are correct, the latter result not attaining conventional significance levels. The picture for conduct resembles that for arithmetic (43%, $z = 1.71$, $p < .10$). (At midyear the children had accurately forecast their mark in 32%, 35%, and 35% of the cases for reading, arithmetic, and conduct respectively.)

For cohort 1 alone, no significant matching was observed while in cohort 2 only reading showed matching significantly greater than chance between the children's year end expectations and marks ($t = 2.89$, $p < .01$).

The discrepance between year-end marks and expectations will be treated in more detail later in this chapter in relation to other variables.

How Do Children's Expectations Change Over the First-Grade Year?

It has already been pointed out that the average level of children's expectations in all three areas increases over the first-grade year. How does this change appear at an individual level? Much of what follows addresses this question by looking at changes in expectations in relation to marks received, to the expectation-mark discrepance at mid-year, to parent-expectations, and the like. The simplest question is to ask how much matching at an individual level there is at midyear and at the end of the year, and this question will be answered first.

About 48% of children hold the same expectations for reading at the end of the year as in the middle of the year, as Table 5.5 shows. While 40% would be expected to show the same expectation by chance, the amount of persistence observed is too large to attribute to chance ($z = 2.18$, $p < .05$). More important, those whose expectations change tend to move up--60% move up while 40% move down--although the trend does not quite attain significance ($\chi^2_1 = 3.18$).

In arithmetic the picture of persistence is more pronounced--49% have the same expectation at midyear and at the end of first grade. Because of the differing margins in arithmetic compared to reading, however, this matching (37%) exceeds chance expectancy by a larger amount, and in fact is highly significant ($z = 3.62$, $p < .01$). There is no significant tendency for expectations in arithmetic to rise or fall over the year, although 56% of those who do change move up.

Table 5.5

		Reading Expectation, Time 2					
		1	2	3	4	Total	Percent
Reading Expectation, Time 1	1	57	33	4		94	54.0
	2	23	23			46	26.4
	3	4	15	2		21	12.1
	4	8	3	1	1	13	7.5
	Total	92	74	7	1	174	
	Percent	52.9	42.5	4 0	0.6		100.0

Table 5.6

		Arithmetic Expectation, Time 2					
		1	2	3	4	Total	Percent
Arithmetic Expectation, Time 1	1	46	25	3	2	76	43.7
	2	21	34	7	1	63	36.2
	3	11	11	5	1	28	16.1
	4	3	1	2	1	7	4.0
	Total	81	71	17	5	174	
	Percent	46.6	40.8	9.8	2.9		100.0

Table 5.7

		Conduct Expectation, Time 2					
		1	2	3	4	Total	Percent
Conduct Expectation, Time, 1	1	48	25	6		79	45.4
	2	27	12	5		44	25.3
	3	13	14	8		35	20.1
	4	7	6	2	1	16	9.2
	Total	95	57	21	1	174	
	Percent	54.6	32.8	12.1	0.6		100.0

Rather surprisingly there is no more correspondence between children's expectations in conduct at the two time points than would be predicted by chance--40% give the same guess on the two occasions. Approximately 36% would be expected to give the same guess by chance. There is, however, a highly significant upward movement of expectations for conduct over the year. About 66% of those children whose expectations change show an increase over the year, a movement that is highly significant ($\chi_1^2 = 10.37$, $p < .01$). A large increase in expectations for conduct marks occurs--about half a grade point--as we have already seen, and this improvement is manifest over the entire marking range.

To sum up: Children's expectations in reading and arithmetic are quite stable over the year and tend to change little. In both cases when there is a change, it tends to be in an upward direction but only the trend for arithmetic is significant. For conduct there is noticeable and significant change overall and it represents a highly significant upward movement.

How Do Expectations Change in Relation to Initial Expectation Level

As Tables 5.5, 5.6, and 5.7 show and as the χ^2-tests for direction of change also document, there is a much stronger tendency for expectations to move up than down. For children with low expectations at midyear (3's or 4's), there is a marked movement upward. For those with expectations of 2 and 3, where movement could occur either up or down, there is no downward movement for reading, and very modest amounts for arithmetic (9 cases out of 51) and conduct (5 cases out of 59). Thus using only data not subject to floor or ceiling effects, the upward thrust appears strong.

How Marks Compare in Different Areas

First Report Card

Marks are assigned in all three areas on a range from 1 (high) to 4. The teachers' marking range is restricted, however. Of 178 cases, no 4's are given in either arithmetic or in reading at the middle of first grade, and only one 4 (in arithmetic) is given at the end. As remarked earlier, average marks are relatively stable over the year in reading and in conduct, but there is a significant increase over the year in the average mark in arithmetic.

Marks in this school are given in relation to the student's supposed ability rather than in relation to others in the class or to

Table 5.8

		Arithmetic Mark, Time 1				
		1	2	3	Total	Percent
Reading Mark, Time 1	1	22	38	1	61	34.3
	2	13	72	12	97	54.5
	3	1	10	9	20	11.2
	Total	36	120	22	178	
	Percent	20.2	67.4	12.4		100.0

Table 5.9

		Arithmetic Mark, Time 2					
		1	2	3	4	Total	Percent
Reading Mark, Time 2	1	43	27	1		71	39.4
	2	23	59	5	1	88	48.9
	3	2	11	7		20	11.1
	4			1		1	0.6
	Total	68	97	14	1	180	
	Percent	37.8	53.9	7.8	0.6		100.0

some absolute standard of achievement. Issues related to the marking practices will be taken up in the Discussion but it is worth noting here that with almost 90% of marks in both reading and arithmetic in the highest two categories as of the first report card, there is little room left for upward movement.

There is greater consistency between marks in these two subjects than would be expected, for 99% of the marks in reading and arithmetic are within one unit of each other. As shown in Table 5.8, significantly more students (57.9%) than would be expected by chance get exactly the same mark in both subjects ($z = 4.11$, $p < .01$). Of the 42.1% who do not get the same mark, 28.1% show a significant tendency ($\chi^2_1 = 9.99$, $p < .01$) for the arithmetic mark to be one unit lower than the reading mark, rather than one unit higher (12.9%). (Elsewhere the tendency of teachers to "mark harder" in arithmetic will be noted.)

The present analysis makes clear that the variance of teachers' marks within children is less than would be expected, even allowing for the restricted range. Given the margins, 12 children would be expected to have marks separated by 2 units; only 2 such children are observed. The numbers are too small, of course, to lead to firm conclusions, but the data are in a direction consistent with a "halo" effect. Teachers are inclined to rate the same child similarly in the different areas.

There is significant matching of marks above chance level between conduct and both reading and arithmetic (48.9%, $z = 2.49$, $p < .01$ for reading; 54.5%, $z = 3.89$, $p < .01$ for arithmetic). In both cases there are about as many children as would be predicted whose conduct and substantive marks are 2 units apart. There is no significant trend for marks in either reading or arithmetic to be higher or lower than marks in conduct.

Marks at End of First Grade

The exact matching of marks in reading and arithmetic increase about 3% from the middle to the end of first grade. Over three-fifths (60.6%) of children (N=180) get the same mark in both subjects at the end of first grade (Table 5.9). This correspondence is highly significant ($z = 5.29$, $p < .01$). The tendency for marks in reading to exceed those in arithmetic that existed earlier disappears by the end of first grade. Teachers display about the same marginal distributions of marks for both subjects (about 40% A's and 50% B's in both) at the end of the year, whereas at midyear arithmetic was marked more stringently (20% A's in arithmetic vs. 34% in reading).

At the end of first grade there is still significant matching between marks at midyear both in conduct and in the two substantive areas (46.7%, $z = 2.02$, $p < .05$ in reading; 48.3%, $z = 2.11$, $p < .05$ in arithmetic). Percentages matching decrease in both areas, however, over the course of the year. These decreases are not significant, but

deserve attention because they are opposite in direction to the original observation of agreement between the two substantive areas. There is no distinct trend for marks to be higher or lower in conduct than in the substantive areas. There is also less clustering around the main diagonal than in the reading or arithmetic comparison, with 11 and 8 children whose reading and arithmetic marks, respectively, are two or more units distant from their conduct mark. This corresponds closely to the midyear data where 9 and 10 children respectively had conduct marks 2 or more units distant from their substantive area marks. This number of children in small, but together with the reduction of exact matching, it probably signifies that teachers make more "fine" distinctions between academic performance and classroom social behavior as the year progresses.

How Expectations Compare in Different Areas

At First Grade, Midyear

Although marks (and thus teachers) signify that performance levels in reading and arithmetic are correlated, children do not anticipate this association. Children who think they will do well in reading at the first report card are apt not to be those who think they will do well in arithmetic.

Guesses before the first report card match in the two areas in only 25% of the cases and by chance 36% matches would be predicted (Table 5.10). This under-matching between reading and arithmetic is highly significant ($z=3.29$, $p < .01$). Children whose expectations in reading and arithmetic do not match expect to do worse in arithmetic than in reading--44% expect to do worse vs. 31% expecting to do better--a difference significant at about the 5% level ($X_1^2 = 3.81$) when only unmatched cases are tested.

A End of First Grade

At the end of first grade significant undermatching is still present ($z = 2.60$, $p < .01$), although more children than earlier (33% vs. 25%) expect the same mark in the two subjects. Chance prediction is 42% (see Table 5.11). In fact the discrepance from chance is roughly of the same magnitude as the discrepance at midyear. More guesses still place arithmetic below reading but, unlike results for midyear, the tendency is not strong enough at the end of the year to attain significance.

In view of the above, conduct marks could not agree well with both arithmetic and reading. At midyear there is a significant excess of matches between reading and conduct (44% match compared to 34% predicted, $z = 2.73$, $p < .01$) and a significant deficit in matches between arithmetic and conduct (25% match compared to 33% predicted,

Table 5.10

		Arithmetic Expectations, Time 1					
		1	2	3	4	Total	Percent
Reading Expectations, Time 1	1	40	49	10	3	102	54.8
	2	27	6	14	1	48	25.8
	3	9	9	1	4	23	12.4
	4	6	3	4		13	7.0
	Total	82	67	29	8	186	
	Percent	44.1	36.0	15.6	4.3		100.0

Table 5.11

		Arithmetic Expectations, Time 2					
		1	2	3	4	Total	Percent
Reading Expectations, Time 2	1	38	47	8	2	95	51.6
	2	45	21	12	2	80	43.5
	3	3	3	1	1	8	4.3
	4	1				1	0.5
	Total	87	71	21	5	184	
	Percent	47.3	38.6	11.4	2.7		100.0

$z = 2.34$, $p < .05$). Children show a significant tendency ($\chi^2 = 5.95$, $p < .01$) to believe they will do worse in conduct than in reading, for 62% of the children with unmatched expectations guess a lower mark in conduct. In arithmetic about as many children believe they will do worse in conduct than in arithmetic as believe they will do better (54% vs. 46%).

At the end of first grade, the significant matching between guesses for reading and for conduct persists (50% of the guesses match compared to 43% expected by chance, $z = 2.03$, $p < .05$). Although the absolute amount of matching has increased over the semester (from 44% to 50%), the margins change in a way to increase the number expected by chance (from 35% to 43%). The "excess" matching at midyear is 9% and at the end of first grade is 7%. In absolute terms then, there is a decrease in over-matching during the time period. There are roughly equal numbers who guess they will do better or worse in conduct compared to reading (48% vs. 52% respectively).

By the end of first grade there are about as many matched guesses between arithmetic and conduct as would be expected by chance (38% are noted and 40% would be expected). Also guesses above and below are roughly equivalent--of the unmatched cases 53% expect to do better in conduct and 47% expect to do better in reading.

To sum up: These children appear to start school with very high expectations overall, but particularly in reading. At first report card 55% expect A's, and although by the end of the year only 52% expect A's, less than 5% expect less than a B. These children are clearly optimistic about the likelihood they will do well in reading and a year's experience in school does nothing but increase their optimism. Hopes in arithmetic are not quite as high, although almost half the children expect an A on the first report card. The most noteworthy difference is that a significant fraction--about one-seventh--look for a C or D in arithmetic at year's end.

What is Impact of First Mark on Expectations Later?

If a child got an A in reading, in 72% of the cases he expects an A at the end of the year. If he got a B, in 54% of the cases he looks for a B again but in almost the entire remainder of the cases, he looks for an A (41%). If he got a C, in only 10% of the cases does he look for a C again; in 45% of the cases he looks for a B and in 45% of the cases he looks for an A!

Can the improvement in forecast accuracy over the year (see preceding section showing more matching between marks and expectations at end-of-year than at midyear) be accounted for in terms of improved accuracy of only those who receive the higher marks? That is, given

Table 5.12

		Reading Mark, Time 2					
		1	2	3	4	Total	Percent
Reading Mark, Time 1	1	42	16			58	33.9
	2	24	55	14		93	54.4
	3	1	14	4	1	20	11.7
	Total	67	85	18	1	171	
	Percent	39.2	49.7	10.5	0.6		100.0

Table 5.13

		Arithmetic Mark, Time 2					
		1	2	3	4	Total	Percent
Arithmetic Mark, Time 1	1	29	4	1		34	19.9
	2	33	78	3	1	115	67.3
	3	3	10	9		22	12.9
	Total	65	92	13	1	171	
	Percent	38.0	53.8	7.6	0.6		100.0

the relative stability of marks between midyear and year end (68% of the children receive identical reading marks both times; 99% receive reading marks less than 2 points away from their midyear mark) it would seem that following anything other than a rule such as "if you got a C this time (midyear) expect a C next time (year end)" would lead to inaccuracy. Since accuracy improves, it might seem that children with C's who expect A's and B's are not the ones providing for any improvement.

If we look at Table 5.12 however, we see that accuracy can be improved by children who get C's. We see that 70% of the children who received a C a mid year in reading get a B at year end. Their marks actually improved which means expecting an improvement should increase accuracy for them. It would be incorrect to conclude that only the children with A's and B's are producing the improvement in grade expectation matching despite the seeming contradiction that children with initially low marks maintain relatively nigh expectations.

One should note that the children who receive C's at midyear are atypical in that the mark they should expect at year end is not the mark they received at midyear, as is the case with children who received A's and B's. The relatively small number of children that receive C's means this distinction can go unnoticed if one simply uses "percent receiving identical marks" in the whole table, as above.

Of the children who received C's at year end, 78% had received a B at midyear. The probability of getting a C after having a B at midyear is smaller than the probability of getting a C after a C at midyear (.15 as compared to .20) but the larger number of cases to which the former applies results in a larger proportion of the year end C's having a midyear B (14 cases) rather than a midyear C (4 cases). In reading only 22% of the children who received a C at midyear received a C at year end. More is said about this differential accuracy in the next section.

In arithmetic the same tendencies appear but are less marked. If a child got an A (N=34), he gets an A again in 85% of the cases. If he got a B, in 4% of the cases he gets a lower mark; in 20% of the cases he gets a higher mark. With a C mark again there is pronounced reason for expecting a higher mark for 67% receive a higher mark.

The picture for conduct is highly similar--of those given C's (N=29), more than 50% improve.

In reading, arithmetic, and conduct, of the children who receive C's at midyear the largest number receive B's at year end (70%, 45%, and 52% respectively). This means that children with the lowest grades should be optimistic if their forecasts are to be accurate. Downward recruitment from the B category is not as pronounced for arithmetic and conduct as for reading and this produeces a decrease in the number of persons who receive C's at the year end compared to mid year.

How Does Children's Expectation-Mark Agreement Vary with the First Mark Received?

In order to discuss tables like Table 5.14, of which many other examples will appear in the remainder of this chapter and in the next chapter, we must first acquaint the reader with the possible patterns of outcomes, and the substantive significance of these patterns. First, of a student receives a mark of 1 his expectation can only be equal to or be lower than the mark. Only values 0, +1, +2, and +3 on the Mark-versus-Expectation variable are possible for such a person. Similarly, with a mark of 4 expectations can only be equal to or higher than the mark. Only -3, -2, -1, and 0 are possible values of the Mark-versus-Expectation variable. For a mark of 2 only -1, 0, +1, and +2 are possible and for a mark of 3 only -2, -1, 0, +1 are possible values.

Second, one must identify the pattern of outcomes expected if children were a) able to judge accurately or test the reality of their performance and adjust their expectations accordingly, or b) acting independently--without any feedback or knowledge--of school performance.

If children formulate their expectations independently of actual performance (marks) the distribution of expectations would be the same for all mark categories. It would, furthermore, be the same as the overall expectation distribution. For example, with an overall expectation distribution of 1 = 55%, 2 = 27%, 3 = 12%, and 4 = 6% we would expect the following pattern.

Mark-Expectation Prediction,
Independent of Performance
(percentages along rows)

		-3	-2	-1	0	+1	+2	+3	
Mark	1				55	27	12	6	100
	2			55	27	12	6		100
	3		55	27	12	6			100
	4	55	27	12	6				100

If, on the other hand, children were adjusting their expectations according to their accurate perceptions of their actual performance the bulk of the cases would fall in the column headed "0" (i.e. no mark-expectation difference) for each of the mark categories, and most of the remaining cases would cluster in the -1 and +1 columns (i.e. small mark-expectation discrepancies for all mark categories). The actual amount of clustering cannot be predicted, of course, from the assumptions given above, so the percentages presented in the following table are hypothetical and made roughly consistent with the previous example for ease of comparison.

Mark-Expectation Prediction,
Assuming Performance Monitoring
(hypothetical percentages along row)

		-3	-2	-1	0	+1	+2	+3	
Mark	1				55	27	12	6	100
	2			19	55	20	6		100
	3		6	20	55	19			100
	4	6	12	27	55				100

Under the "chance" or "marginal probability" assumption the bulk of the cases fall on the diagonal that rises from the lower left corner of the table and a declining set of percentages equivalent to the marginal percentages is expected in the column headed "0" discrepance. Under the "reality testing" assumption the bulk of the cases fall in the column headed "0" discrepance. Note that the two assumptions give similar shaped distributions in the row headed "Mark = 1." In any given case however, it may be possible to decide between the reasonableness of the two assumptions by looking at the magnitude of the percent in the row = 1, column = 0 cell. If, for example, 95% of cases appear in this cell the data are not inconsistent with "reality testing" assumption which only claimed the bulk of the cases would be found in this cell. Such a finding would be inconsistent with the "marginal probability" assumption which specified the magnitude of this percent as approximating the overall percent of expected (i.e. in our example 55%).

With these two assumptions and their resultant models in mind let us examine Table 5.14 which presents the data for reading marks and expectations at midyear. The distribution of reading expectations at midyear was 1's = 55%, 2's = 27%, 3's = 12% and 4's = 6%. The reader will recognize these as the percents used in the "hypothetical" example above which can be used for a direct comparison. The data clearly tend to follow the "marginal probability" model. The bulk of the cases, for the 3 marks the teachers assign, fall on a diagonal decreasing to the left of the Mark = 1, Mark-Expectation = 0 cell. Furthermore, all these cells (51%, 58%, and 56%) approximate the 55% expected on the basis of our overall percent of 1's expected. The data do not perfectly fit the "marginal probability" model but approximate it more closely than the "reality testing" model (remembering the mark = 1 distribution is consistent with both).

It appears, then, that reading performance (as judged by teachers) is not information that children incorporate into their midyear reading expectation. This appears to be specifically so for those children who receive 2's and 3's, and possibly so for those who receive 1's although here it is hard for the model to be wrong in the case of 1's.

Table 5.14

		Expectation-Mark, Reading, Time 1								
		-3	-2	-1	0	+1	+2	+3	Total	Percent
Reading Mark, Time 1	1				51% 30	39% 23	5% 3	5% 3	59	100%
	2			58% 55	22% 21	14% 13	6% 6		95	100%
	3		56% 10	11% 2	22% 4	11% 2			18	100%
	4									
	Total		10	57	55	38	9	3	172	

By noting that those children with low marks have larger proportions of extreme and moderate discrepancies between their expectations and marks, one might take Table 5.14 to show that children with the poorest marks are also the children who are poorest at forecasting their marks. The discussion of the "marginal probability" model which fit the data, however, suggests that all the children may be using an equivalent forecasting scheme which is independent of performance. Children who do better may be better at forecasting than children who do poorly but at this stage of the analysis it seems prudent to explain the finding on the basis of the simplest model which fits the data.

Arithmetic

For arithmetic the overall distribution of expectations is: 1's = 44%, 2's = 37%, 3's = 14%, and 4's = 5%. As with reading the bulk of the cases seem to fall along a diagonal descending to the left of the mark = 1, discrepance = 0 cell, again supporting the "marginal probability" model. The majority of the cases clearly do not fall in the "0" discrepance column (which approximate the marginal distribution as the marginal probability would predict), but the percentages in the diagonal do not fit perfectly either. Most notably, the distribution in the mark = 3 row show the bulk of the cases occurring between the positions our competing models would predict. This may indicate that some low performance children are accurately judging their performance in arithmetic and lowering their expectations slightly (but not sufficiently to remove discrepance), or it simply may be an artifact due to the small number of cases (20) involved. (See Table 5.15)

In arithmetic, then, most of the evidence is consistent with an observation that children are not using their performance in arithmetic (as it is viewed by the teachers) in formulating their expectations.

Conduct

For conduct the overall distribution of expectations is: 1's = 46%, 2's = 26%, 3's = 19%, and 4's = 9%. The pattern is rather similar to that for arithmetic. The bulk of the cases fall on the diagonal (not the "0" column), supporting the non-reality testing model, but the diagonal percentages (58%, 41%, 39%) do not reproduce as closely as one might wish the expected percentage (46%). In this case it is the over-representation in the mark = 1, discrepance = 0 cell that seems aberrant. As mentioned earlier, if the percentage in this cell significantly exceeds the percent expected (the overall percent receiving 1's), evidence is thereby consistent with a reality testing model. Of the five cells on the diagonal or in the "0" column (neglecting the 4th row due to insufficient cases) four are consistent with the non-reality testing model (41% vs. 46%; 25% vs. 26%; 39% vs. 46%; and 21% vs. 19%) and one may be more consistent with a reality testing model (59% vs. 46% expected by non-reality testing). (See Table 5.16)

Table 5.15

Expectation-Mark, Arithmetic, Time 1

		-3	-2	-1	0	+1	+2	+3	Total	Percent
Arithmetic Mark, Time 1	1				49% 17	40% 14	11% 4		35	100%
	2			45% 53	35% 41	16% 18	4% 5		117	100%
	3		25% 5	45% 9	15% 3	15% 3			20	100%
	4									
Total			5	62	61	35	9		172	

Table 5.16

Expectation-Mark, Conduct, Time 1

		-3	-2	-1	0	+1	+2	+3	Total	Percent
Conduct Mark, Time 1	1				58% 33	25% 14	14% 8	3% 2	57	100%
	2			41% 35	25% 21	22% 19	12% 10		85	100%
	3		39% 11	32% 9	21% 6	7% 2			28	100%
	4	50% 1			50% 1				2	100%
Total		1	11	44	61	35	18	2	172	

In summary, most of the data on reading, arithmetic, and conduct support the statement that children of all ability levels may formulate their expectations the same way before the first report card is received, in a way independent of performance as viewed by teachers.

How Does Children's Expectation-Mark Agreement Vary with the Mark Received at the End of First Grade?

By the end of first grade there is a noticeable movement toward a pattern of expectation-mark discrepancies consistent with the reality-testing model for reading for the students who get high marks. There is both the predicted piling up in the Row = 1, Column = 0 cell, and a marked shift in the Row = 2 distribution. For the Row = 3 children, results are mixed. It appears that optimism has moderated, but guesses about performance are still biased strongly in a positive direction. See Table 5.17.

For arithmetic (Table 5.18) there appears again a movement toward a pattern consistent with a reality-testing model. There is a slight piling up in the "0" column--an excess for Row = 1 and a mode for Row = 2. Again the children in the in the Row = 3 category seem aberrant and inclined to maintain the pattern manifest at midyear but the number of children (13) in this row is small so strong conclusions are not warranted.

In conduct, unlike reading and arithmetic, there is the potential for a poor grade. Five children (2%) actually receive the lowest grade and 15% get C's or D's, a total of 17%, whereas 11% in reading and 8% in arithmetic get the lower two grades.

Despite the fact that conduct should be more capable of monitoring by the child, it only shows a slight tendency toward reality-testing (see Table 5.19). The mark = 1, discrepance = 0 cell contains more cases (63%) than would be expected on the basis of the "marginal probability" model (52%) and the mark = 3 row has its mode shifted one column away from when the marginal probability model would predict it to be (i.e. on the diagonal). Children who receive a mark of 2, however, exhibit almost exactly the pattern the marginal probability model would predict (50% vs. 52%, 36% vs. 35%, 13% vs. 12%, and 1% vs. 1%).

In sum: There seems initially, before the first report, a considerable push for expectations to be high and for expectations to be similar regardless of the child's actual level of performance at midyear. By the end of the second semester a noticeable though not extreme shift toward reality-testing has occurred for reading. Arithmetic and conduct, however, show only very slight departures from the "marginal probability" model. These departures only minimally exceed those observed for arithmetic and conduct at midyear. At year end expectations remain largely independent of actual performance in arithmetic and conduct while a slight dependence is exhibited in reading.

Table 5.17

Expectation-Mark, Reading, Time 2

		-3	-2	-1	0	+1	+2	+3	Total	Percent
Reading Mark, Time 2	1				68% 47	29% 20	1% 1	1% 1	69	100%
	2			43% 37	51% 44	6% 5			86	100%
	3		32% 6	63% 12	5% 1				19	100%
	4		100% 1						1	100%
Total			7	49	92	25	1	1	175	

Table 5.18

Expectation-Mark, Arithmetic, Time 2

		-3	-2	-1	0	+1	+2	+3	Total	Percent
Arithmetic Mark, Time 2	1				57% 37	38% 25	5% 3		65	100%
	2			41% 39	43% 41	14% 13	3% 3		96	100%
	3		38% 5	31% 4	23% 3	8% 1			13	100%
	4		100% 1						1	100%
Total			6	43	81	39	6		175	

Table 5.19

Expectation-Mark, Conduct, Time 2

		-3	-2	-1	0	+1	+2	+3	Total	Percent
Conduct Mark, Time 2	1				63% 40	30% 19	6% 4		63	100%
	2			50% 43	36% 31	13% 11	1% 1		86	100%
	3		29% 6	43% 9	24% 5	5% 1			21	100%
	4	40% 2	40% 2	20% 1					5	100%
	Total	2	8	53	76	31	5		175	

All children seem very optimistic before they get a first report card and optimistic in much the same way. Children whose grades are already at or near the ceiling cannot be more optimistic than their performance and this leads to a confounding of effects that must be acknowledged in interpreting the findings.

How Does Children's Expectation-Mark Agreement Vary with Expectation Level?

Reading

As already noted there is a marked optimism seen in children's initial expectations in reading for 55% look for A's on the first report card (34% receive one) and 82% look for an A or a B. This overall optimism, however, disguises some interesting pessimism when expectation-mark agreement is broken down by expectation level (see Table 5.20). (The reader should note that by examining diagonals in this table, one can locate persons who received a particular mark. Thus reading down the right-most marginal, 30, 23, 3, 3, one has a summary of those children, 59, who received an "A" in reading on the first report card.)

The majority of children actually receiving an A expected one (30/59 or 51%) but many more children expecting A's did not receive one (65/95 or 68%). The majority of extraordinarily high hopes were therefore not confirmed. On the other hand, half the children expecting a "B" (23/46 or 50%) actually received an "A," so children expecting B's are generally under- rather than over-optimistic. The under-optimism is even more noticeable among those children who expect C's or D's--only 13% of the total (4/31) actually receive a C and none receives a D. Altogether 32% of children forecast their mark exactly, and most of these get an A. The general trend seems therefore to be one of inaccuracy--the overall mean suggests an inaccuracy based on over-optimism but further analysis suggests both over- and under-optimism.

The presence of both over- and under-optimism, while worth noting because of its substantive impact, is not entirely unexpected. If one disregards exact matching of marks and expectations, it is easy to see that children expecting a 4 can only underestimate. Likewise a child expecting a 1 can only overestimate (or "exactly" forecast his mark). Those expecting 2's and 3's can either over- or underestimate but these categories can lead to extreme discrepancies occurring in only a single direction. Therefore, given some children expecting 3's and 4's, the presence of some underestimation is guaranteed.

One can ask whether children with naturally high expectations are superior (or inferior) at providing accurate forecasts. If children of all expectation levels showed no accuracy at all in forecasting their marks each expectation level would be assigned a mark distribution corresponding to the overall mark distribution (i.e. 1's = 34%, 2's = 55%, and 3's = 11%). This leads to the table of expected values shown below.

Expected Values for Complete Lack of
Forecast Accuracy

		Expectation-Mark, Reading, T1							
		-3	-2	-1	0	+1	+2	+3	
Child's Reading Expectation, T1	1		11	55	34				100%
	2			11	55	34			100%
	3				11	55	34		100%
	4					11	55	34	100%

(The reader should note the parallel between this analysis and that done previously for children who received various marks.) A comparison of this table with Table 5.20 indicates a reasonable similarity, indicating lack of a superior forecast ability on the part of children at any of the expectation levels. Indeed, the most noticeable discrepance in terms of difference between percentages and the case base for the percentages occurs for the children who expect a 2 and actually get a 1. This cell is over-represented indicating these children are doing worse at forecasting than if their forecast bore no relation to their mark! As a general statement, however, forecast ability does not seem to consistently vary with expectation level and the accuracy of forecasting seems to be at about chance levels. While the above discussion is couched in terms of "forecasting ability" the reader should note it could just as reasonably have been discussed in terms of the causal efficacy of expectations. If expectations highly influenced marks, or even if different expectation levels had different effects on marks, deviation from the "chance" model would have been expected. That is, the mark distributions within each expectation level should have deviated from the marginal distribution. In particular the column headed "zero discrepance" and those with small discrepancies (+1 or -1) should have been over-represented if expectations were displaying substantial causal efficacy.

Arithmetic

In arithmetic optimism is considerably less--only 44% look for an A (and of these only one in four receive an A). Overall, however, the high percentage looking for an A or a B (82%) resembles the picture for reading (82%). Of those looking for a B in arithmetic, however, the vast majority (41/64 or 64%) receive one, a situation which differs from that noted in reading where most received A's. Again one notes a degree of over-pessimism because of those looking for a C or a D--19%--only a few persons, 3 out of 33, receive even as low a mark as a C. The data are shown in Table 5.21.

As with reading we can ask for arithmetic whether children of different expectation levels display differential forecasting ability. Since the "marginal" distribution of marks is 12% - 3's, 68% - 2's, and

Table 5.20

Expectations-Mark, Reading, Time 1

		-3	-2	-1	0	+1	+2	+3	Total	Percent
Reading Expectation, Time 1	1		11% 10	58% 55	32% 30				95	100%
	2			4% 2	46% 21	50% 23			46	100%
	3				20% 4	65% 13	15% 3		20	100%
	4					18% 2	55% 6	27% 3	11	100%
Total			10	57	55	38	9	3	172	

Table 5.21

Expectation-Mark, Arithmetic, Time 1

		-3	-2	-1	0	+1	+2	+3	Total	Percent
Arithmetic Expectation, Time 1	1		7% 5	71% 53	23% 17				75	100%
	2			14% 9	64% 41	22% 14			64	100%
	3				12% 3	72% 18	16% 4		25	100%
	4					38% 3	63% 5		8	100%
Total			5	62	61	35	9		172	

Table 5.22

Expectation-Mark, Conduct, Time 1

		-3	-2	-1	0	+1	+2	+3	Total	Percent
Conduct Expectation, Time 1	1	1% 1	14% 11	44% 35	41% 33				80	100%
	2			20% 9	48% 21	32% 14			44	100%
	3				18% 6	58% 19	24% 8		33	100%
	4				7% 1	13% 2	67% 10	13% 2	15	100%
	Total	1	11	44	61	35	18	2	172	

20% - 1's, this is the distribution we would expect in the rows of the table if children of all expectation levels displayed "chance" levels of forecasting ability. Table 5.21 very closely approximates this (noting the small case base for an expectation of 4). The data do not contradict a statement that children of all ability levels forecast at the chance level or, equivalently, that children's expectations are not differentially influencing the mark distributions.

Conduct

In comparison with the previous tables for reading and arithmetic two factors stand out in Table 5.22 which summarizes data for conduct. The first is the reduction in optimism--28% look for C's or D's. The second is the wide range in discrepancies--almost 19% of the guesses are off by 2 or more units in either direction. This lack of accuracy is paradoxical in a sense because conduct is the area in which the child presumably has the most clear-cut feedback and in which he has most control over performance.

Again to answer the question of whether or not children of different expectation levels display differential forecasting ability we must start from the marginal mark distribution: in this case 1% receive 4's, 16% - 3's, 49% - 2's, and 33% -1's. Comparing Table 5.22 with the "chance expectation" table arrived at by placing this distribution in the appropriate portions of the rows of a table, shows again that the chance model fits quite well. Thus children of all expectation levels (with the possible exception of those expecting 1's) are forecasting their marks at near chance levels.

In summary, in all three areas both over and under optimism is present--the over optimism naturally occurs to children with high expectations and under optimism occurs to those with low expectations. In all three areas children of all expectation levels show only a chance level of forecasting their marks or, equivalently, in all three areas the children's differential expectations are not differentially influencing to mark distributions at midyear (i.e. the midyear mark distribution is the same no matter what level of children's expectations is considered). This lack of a midyear effect contrasts with that noted in the next section concerning both midyear and year end observations.

The Effects of Feedback

Reading: Feedback Effect on Marks

About 23% of the marks in reading went up between the middle and end of the first grade, 58% remained stable, and 19% declined (based on a sample of 165 children). With one exception all movement was confined to a single step (i.e., from B to A, C to B, etc.).

Table 5.23

		Change in Reading Mark, T1 to T2			
		Up	Same	Down	
Child Previously Did B,S,W Than He Expected (T1)	Better	4	32	12	48
	Same	10	32	9	51
	Worse	24	32	10	66
		38	96	31	165

Table 5.24

		Predicted Reading Mark Change, T1 to T2			
		Up	Same	Down	
Child Previously Did B,S,W Than He Expected (T1)	Better	11.05	27.93	9.02	48
	Same	11.75	29.67	9.58	51
	Worse	15.20	38.40	12.40	66
		38	96	31	165

If a child's year end mark improved over his midyear mark (N = 38), in 63% of the cases his mark at the middle of the year had been less than he expected. If a child's marks deteriorated (N = 31) in 39% of cases his mark was better than he had expected. The 3 X 3 table, Table 5.23, is a collapsed version of a 7 X 7 table that relates all possible types of mark changes to all possible types of mark-expectation relationships at midyear. A χ^2 test on the four corners of the table shows the effect of expectations upon later improvement (or deterioration) in marks to be highly significant ($X_1^2 = 9.18$, $p < .01$) given an initial discrepance between mark and expectation. There is a significant tendency for marks and expectations to move toward consistency. But a more interesting comparison is one that can be applied to the entire table and one which has a causal frame of reference. In terms of causation, expectations being higher than the mark should lead marks to rise; expectations being the same as the mark should lead to no change in mark; expectations being lower than the mark should lead the mark to fall. In a perfect causal system in which expectations were the only causes, all cases would fall on the "minor diagonal" of the table above if consistent with the pattern just outlined.

With the margins as given the expected values are those given in Table 5.24. The cells on the minor diagonal (enclosed in rectangle) sum to 53.89, and their sum departs from the observed data in the direction predicted by the causal hypothesis--9 more children (24-15) have marks that rise, 3 more children have marks that fall, and 2 more have marks that stay the same. The easiest way to assess the "minor diagonal" effect is by the same statistic used earlier to assess the major diagonal (match) effect. Such a test here reveals a significant effect ($z = 2.43$, $p < .01$). The magnitude of this effect is hard to assess but in that column representing those whose marks improve one can see that most (63%) are recruited from students who did not do as well as they expected earlier.

Reading: Feedback Effect on Expectations

Marks are affected by an earlier mark-expectation discrepance but the inertia of marks seems relatively large--there is a strong tendency for marks to stay the same no matter what the discrepance. Expectations, though, are more responsive in an optimistic direction, for of those (47) whose mark was higher than expected, 38 (81%) raise their hopes. Expectations at the end of the year are more responsive to earlier marks than are the marks at year end. This is hardly surprising since a child's expectations are something he himself controls, whereas his mark depends on someone else's behavior.

If a child receives a mark higher than he expected in reading at the end of the first semester (as 28% of 166 children do), then his response is very likely (chances are better than 4:1) to increase his expectations at the end of the second semester. If he gets exactly what he expected at the end of the first semester (as 33% do) then he will be likely to keep that expectation (chances are 3 out of 5);

Table 5.25

		Change in Reading Expectations, T1 to T2			
		Up	Same	Down	
Child Previously Did B,S,W Than He Expected (T1)	Better	38	9	0	47
	Same	12	35	7	54
	Worse	0	36	29	65
		50	80	36	166

but if he does change, he is about twice as likely to increase his hopes as to decrease them. If he gets a lower mark than he hoped for (as 39% do), then he will not increase his expectations at the end of the second semester but he is somewhat more likely to hold to his original expectation than to lower his sights (chances are 6 out of 11).

This analysis based upon combined cohorts presents a more complete picture of the response of expectations to previous marks than is available in either cohort and documents with more certainty the trends seen there.

The matter can be explored most easily by condensing a 7 X 7 table of discrepancies vs. changes in expectations to a 3 X 3 table, Table 5.25. First, the degree of matching above chance is highly significant ($z = 8.05$, $p < .01$) and amounts to 61%. Thus if a child received a higher mark than he expected, his hopes tend to rise. If he got what he expected, his hopes tend to remain stable. If he got less than he expected, his hopes tend to decrease. But, and this is perhaps the most important part of the table, if he got less than he hoped, his hopes are more likely to remain fixed than to drop (36 vs. 29 children). That is, there is some tendency for these children's expectations to decline, but this chance of decline is still less than even. If he got more than he hoped, on the other hand, his hopes respond vigorously (38 vs. 9 children) and the odds for an increase are over 4:1. If he got exactly what he expected his hopes may change in either direction with perhaps some bias in the positive direction. This receptivity to good news and relative resistence to bad news can be termed a "bouyancy effect."

Arithmetic: Feedback Effect on Marks

About 27% of children's marks in arithmetic increase from the middle to the end of the first grade year, and less than 5% decrease. (In reading 23% increased and 19% decreased.) The number of cases decreasing in arithmetic is too small to justify analysis.

For children whose marks increase, 51% had received a mark <u>less</u> than they hoped for earlier (compared to 36% who received what they expected and 13% who received more). Of those whose marks improve, then, the majority had hoped for a higher mark than they received. As before, a Z-value based on the minor diagonal can be used to test the hypothesis that marks and expectations move in a way to improve their consistency, and the test is borderline ($z = 1.52$, $p < .10$, one-sided). The magnitude of the movement is less for arithmetic than for reading.

Arithmetic: Feedback Effect on Expectations

Again, as was true for reading, children's expectations are more labile than marks in terms of the earlier (T1) expectation-mark discrepance--some children (over 10%) raise or lower their expectations by 2 or 3 units. Marks have a much narrower range of variation.

Table 5.26

		Change in Arithmetic Mark, T1 to T2			
		Up	Same	Down	
Child Previously Did B,S,W Than He Expected (T1)	Better	6	32	4	42
	Same	16	41	2	59
	Worse	23	39	2	64
		45	112	8	165

Table 5.27

		Change in Arithmetic Expectations, T1 to T2			
		Up	Same	Down	
Child Previously Did B,S,W Than He Expected (T1)	Better	30	11	1	42
	Same	13	35	11	59
	Worse	2	36	27	65
		45	82	39	166

The easiest way to see overall trends in expectation change as a function of marks received, is in a 3 X 3 table, Table 5.27 (summarized from a 7 X 7 table). Again the amount of matching (55%) is large and highly significant ($z = 6.08$, $p < .01$) when compared to the amount expected by chance (34%). Also as earlier was true for reading, the majority (58%) of children who received negative feedback (did worse than expected) at least maintain their expectations at the same level despite the receipt of a mark lower than they expected. Also, as was true earlier for reading, there is a vigorous response to positive feedback--71% respond with raised hopes. If the child previously got exactly what he expected, about the same number of children raise hopes as lower them (13 vs. 11).

With positive feedback, then, the majority response is to raise hopes. With negative feedback, a considerable fraction lower hopes, but a larger fraction maintain hopes at the same level. In sum, a bouyancy effect is observed for arithmetic.

Conduct: Feedback Effects on Marks

In conduct there is a positive shift in marks given at the end of the year--22% of children get higher marks while 11% decline. Again trends are clarified by a 3 X 3 summary of the midyear discrepance between expectations and marks and mark changes (Table 5.28). There is a significant over-representation of cases on the minor diagonal ($z = 2.57$, $p < .01$). The amount of matching is 41.8%, or about 8% more than chance prediction. There are about equal numbers of children with positive and negative discrepances (earlier in reading and arithmetic there were more children with positive discrepances, i.e., expectations higher than marks).

Since there are some children whose marks decline (11%), one can also examine the effect of receiving a mark higher than one expected. If a child's mark was higher than he hoped in conduct, his mark is less likely to shift down (15% shifted down) than to remain stable (72% remain stable) but the probability of his mark shifting down is slightly greater than chance would predict. This picture replicates what happened for arithmetic.

The distribution of changes in conduct marks is almost identical for the neutral and positive feedback cases--8, 40, 7 compared to 7, 39, 8. With neutral and positive feedback, marks show a slightly but not significantly higher-than-chance probability of staying the same (40 vs. 37 expected, 39 vs. 36 expected, respectively). Only if the discrepance is one where a child did worse than he expected is there a noticeable effect later. In this case marks tend to increase significantly ($X_1^2 = 12.04$, $p < .01$). Since 21 cases display an increase out of 56, the effect is also large in practical terms. Conduct would be an area where a child could increase the quality of his performance more readily than in arithmetic, say, and it is therefore sensible for the effect in conduct to be larger.

Table 5.28

		Change in Conduct Mark, T1 to T2			
		Up	Same	Down	
Child Previously Did B,S,W Than He Expected (T1)	Better	7	39	8	54
	Same	8	40	7	55
	Worse	21	32	3	56
		36	111	18	165

Table 5.29

		Change in Conduct Expectations, T1 to T2			
		Up	Same	Down	
Child Previously Did B,S,W Than He Expected (T1)	Better	47	5	1	53
	Same	15	33	10	58
	Worse	3	28	24	55
		65	66	35	166

Conduct: Feedback Effect on Expectations

The expectation changes for conduct from the middle to the end of the first-grade year resemble the picture for similar change in reading and arithmetic: much more lability characterizes changes in expectations than in marks. Again a 3 X 3 matrix makes changes easier to see (Table 5.29). Almost equal numbers of children expected more, the same, or less than the mark they received, but of those who received more than they expected there is a strong tendency (89%) to subsequently alter expectations up, in the direction of consistency. If a child got exactly what he expected, there is movement both up and down, with no particular trend in either direction. But if a child got less than he expected, as before with reading and arithmetic, his expectations are more likely to remain stable than to decrease (56% do not decline). His response to negative feedback in the direction of consistency is much smaller than his response to positive feedback. A buoyancy effect is observed here also.

There is a highly significant amount of over-matching ($z = 8.03$, $p < .01$) and the absolute level--63%--is impressive. In other words, for conduct if a child earlier has received a mark he expected his expectations tend to remain stationary. If he earlier received more than he expected, his expectations tend to rise, and conversely if he received less than he expected, his hopes tend to fall, but as already remarked, the tendency to fall is much less than the tendency to rise.

Summary of Feedback Effects

Effects on Marks

In all cases the majority of children received the same mark at the end of first grade as at midyear of first grade. That is, for reading, arithmetic, and conduct generally and for all the breakdowns of each of these into distributions of students who did better, the same, or worse than they expected, the majority in any breakdown can expect to get the same mark at year end as at midyear.

There is also, however, a consistent (and significant in the case of reading and conduct) pattern for cases on the minor diagonal to appear with greater than chance expectations. More specifically, there is a consistent observation of the following types of cases occuring more often than would be expected by chance:

a) A child's marks going up at year end and after the child did worse than he expected at midyear.

b) A child's mark remaining the same at year end after the child got the same mark he expected at midyear.

c) A child's mark going down at year end after the child did better than he expected at midyear.

These observations are consistent with the notion that student expectations causally affect the marks students receive, even if the degree of association is rather small. As noted earlier, the restriction in range on the marking scale and the similar margins at mid- and end-of-year mitigate against effects being displayed (The power of tests is low).

Effects on Expectations

In all three areas (reading, arithmetic, and conduct) there is a significant tendency for children's expectations to move toward being consistent with the mark they received at midyear. If the midyear mark was above the children's expectations, their expectations moved up substantially; if the midyear mark equalled the children's expectations, their expectations showed less change than would be expected by chance; and if the midyear mark was below the children's expectations, their expectations showed a greater than chance tendency to decline but in all three cases (reading, arithmetic, and conduct) this decline was not as pronounced as the previous two effects (the majority of children maintained their expectations at the same level it had been at midyear).

When information on marks is combined with that on expectations, it looks as though when a child gets less than he hoped he does not lower his sights--his expectations remain the same and there is a tendency to bring marks in line. If he got more than he hoped, his expectations immediately rise and his mark stays the same. There is altogether what might be called a bouyancy effect--a tendency for marks <u>or</u> expectations to <u>rise</u> to narrow the discrepance between them. If the child does better than he expected his hopes rise; if he did worse than he expected his marks rise. If he gets good news he is receptive, whereas he seems relatively impervious to bad news. (The strongest trends observed in the data are stated without qualification in the above summary--i.e., without regard for weak reverse trends--for simplicity of presentation.)

Children's Ability to Forecast

How does discrepance at midyear compare with the end-of year discrepance in the mark-expectation relation?

Reading

Children get better at anticipating their marks, particularly at <u>not underestimating</u>. A summary of a larger table showing how expectations in reading relate to marks in reading (higher, same, or lower) at midyear and at the end of first grade appears in Table 5.30. The percentage agreeing jumps up noticeably from 31% to 55%. To test the significance of this increase in agreement, note that 36 cases remain the same, 14 cases move away from agreement and 52 cases move toward agreement. The χ^2 test for related samples gives a value of 20.74, significant far beyond the .01 level, indicating that the 24% increase in agreement is highly significant.

Table 5.30
Mark-Expectation Discrepance for Reading, T1 vs. T2

		Child Expected to Do B,S,W Than He Did in Reading			
		First Semester			
		Better	Same	Worse	
Second Semester	Better	30	8	13	51
	Same	30	36	22	88
	Worse	5	6	11	22
		65	50	46	161

Table 5.31
Mark-Expectation Discrepance for Arithmetic, T1 vs. T2

		Child Expected to Do B,S,W Than He Did in Arithmetic			
		First Semester			
		Better	Same	Worse	
Second Semester	Better	23	8	10	41
	Same	29	34	14	77
	Worse	11	16	16	43
		63	58	40	161

The result is also significant in practical terms for the majority of children move toward agreement.

The previous pessimists and the previous optimists are about as likely to be recruited into the "agree-ers" category--46% of those who used to have too-high hopes now have their expectations in line and 48% of those who used to have too-low hopes are now in line. More previous optimists are recruited (30) than pessimists (22) simply because there were more optimists to be recruited from (65 compared to 46). On overshooting, there is a slight tendency for those who had hopes too low to move further up than the reverse--28% of those with too-low hopes earlier vs. 8% of those with too-high hopes earlier modify their expectations or their marks to such an extent that the discrepance between marks and expectations has now reversed its direction.

A tendency toward optimism exists at the end of the year, as the margins of Table 5.30 show, and the tendency strengthens when the increased matching is allowed for. That is, in terms of absolute numbers of children being overconfident, 40% at midyear and 32% at the end of first grade expect a mark higher than they get. But it is more pertinent to see how those not matching divide up at both times. Of 111 not matching at midyear, 59% have expectations too high. Of 73 not matching at the end of first grade, 70% have expectations too high.

Arithmetic

A similar analysis to that just give for reading can also be given for arithmetic. The first step is again to condense a larger table to a 3 X 3 table (Table 5.31). Again the amount of matching improves, from 36% to 48%, and a chi-square test ($X_1^2 = 4.83$, $p < .05$) shows a significant movement toward agreement: The movement toward agreement in arithmetic is not as strong as it was for reading. A test of the significance of differences in movement (24% vs. 12%) falls just short of the 5% of significance level ($X_1^2 = 3.53$).

From another standpoint also the changes in agreement between expectations and marks at the two time points for reading and for arithmetic differ. At the end of first grade, 48% are matched but there is not the over-optimism in arithmetic that there is in reading. Of those disagreeing (N = 84) about the same number overestimate as underestimate, 49% vs. 51%. The tendency toward over-optimism is reduced rather than augmented.

These findings must be viewed in the light of teachers' marginal distributions of marks in arithmetic. As noted elsewhere, teachers marked "easier" at the end of first grade than in the middle, so it may be proper to attribute the apparent equality in over- and under-optimism on the part of students to the shift in the teachers' marginal distributions of marks. If the teachers had assigned approximately the same marginals in arithmetic at the two time points (as they in fact did in reading), then presumably more children would have received a lower mark than they expected, leading to over-optimism.

Table 5.32
Mark-Expectation Discrepance for Conduct, T1 vs. T2

		Child Expected to Do B,S,W Than He Did in Conduct First Semester			
		Better	Same	Worse	
Second Semester	Better	26	18	14	58
	Same	22	23	25	70
	Worse	7	12	14	33
		55	53	53	161

Conduct

A similar kind of analysis for mark-expectation discrepance is given for conduct in Table 5.32. First, does agreement between expectations and marks increase? The agreement at midyear is 33% and that at the end of first grade is 43%. This is the smallest increase in agreement noted so far, and is the first not to attain significance (X_1^2 = 3.32, N.S.). There is, however, as was true for reading, relatively more over-optimism at the end of the year than there was in the middle. Of those who do not manifest agreement (N = 91), 64% are over-optimistic vs. 51% earlier. There is some improvement in teachers' marginals for conduct over the year (83% received A's or B's at midyear compared to 85% at the end of the year) but the improvement is slight. Therefore in conduct the main findings are (1) the failure for agreement to increase significantly and (2) a relative increase in over-optimism with time (like that seen for reading).

To sum up: Improvements in children's ability to forecast are hard to assess because if teachers' marking practices differ from one report card to the next, the target of the forecasts is shifting. For reading, where teachers are consistent, there is a large increase in agreement between marks and expectations in both statistical and practical terms. For those whose expectations and marks are not in line at year end, optimism far exceeds pessimism. For arithmetic, where teachers mark more easily at year's end, there is a movement toward agreement which is less strong than the movement in reading. Any over-optimism is probably masked by the teacher shifts. For conduct there is the smallest movement toward agreement coupled with a tendency toward over-optimism. Children do improve in forecasting, then, but not equally across the three preformance areas.

Parents' Expectations Compared to Children's Expectations

In both cohorts of middle class children, parents are optimistic but guarded. As noted in Table 5.1, parents' expectations are fairly stable across areas, being 1.89, 1.93, and 1.93 for reading, arithmetic, and conduct for the combined cohorts. A large majority (64%) expect a "2" in reading, a "2" in conduct (64%), and 59% expect a "2" in arithmetic.

Cohort 1 children exceeded those of cohort 2 in optimism--significantly so in arithmetic and conduct ($t = 2.67$, $p < .01$ and $t = 3.19$, $p < .01$ respectively), not significantly in reading ($t = 1.26$). Only cohort 1 children are noticeably more optimistic in their expectations than parents ($t = 3.78$, $p < .01$; $t = 1.89$, $p < .05$; $t = 2.38$, $p < .01$ for reading, arithmetic and conduct respectively for cohort 1--all cohort 2 t values are insignificant). Over 55% of the children in both cohorts taken together expect the highest grades in reading and 46% and 45% expect A's in arithmetic and conduct respectively.

The matching between expectations of parents and children is about what would be predicted by chance, the observed percentages matching

Table 5.33

Child's Reading Expectation, Time 1

		1	2	3	4	Total	Percent
Parent Reading Expectation, Time 1	1	17	15	4	2	38	24.8
	2	57	24	12	5	98	64.1
	3	11		2	2	15	9.8
	4			1	1	2	1.3
	Total	85	39	19	10	153	
	Percent	55.6	25.5	12.4	6.5		100.0

Table 5.34

Child's Arithmetic Expectation, Time 1

		1	2	3	4	Total	Percent
Parent Arithmetic Expectation, Time 1	1	23	12	2	1	38	24.8
	2	40	30	16	4	90	58.8
	3	8	12	3		23	15.0
	4				2	2	1.3
	Total	71	54	21	7	153	
	Percent	46.4	35.3	13.7	4.6		100.0

Table 5.35

		Child's Conduct Expectation, Time 1					
		1	2	3	4	Total	Percent
Parent Conduct Expectation, Time 1	1	14	10	7	1	32	21.6
	2	42	26	18	9	95	64.2
	3	10	4	3	2	19	12.8
	4	1		1		2	1.4
	Total	67	40	29	12	148	
	Percent	45.3	27.0	19.6	8.1		100.0

being 28.8% in reading, 37.9% in arithmetic, and 29.1% in conduct (N = 153, 153, and 148 pairs respectively) (Tables 5.33, 5.34, 5.35). Furthermore there is little clustering around the main diagonals in the tables. Clustering would indicate closeness, if not exact, matching. Case by case, there is little correspondence between what parents expect and what children expect.

Discrepances between Parents' Initial Expectations and the First Report Card

There is little correspondence (no significant matching) between what children expect and the marks they receive on first report card in any of the three areas (see Tables 5.2, 5.3, 5.4). There is, however, a highly significant ($z = 2.80$, $p < .01$) amount of matching between parents' expectations and children's marks in reading at the first report card. The accuracy in matching (55%) exceeds chance (45%) by a 10% margin (Table 5.36). In other words, there is more matching than chance would predict. The same situation prevails in arithmetic (56% matches vs. 47% expected by chance; $z = 3.00$, $p < .01$) and in conduct (54% matches vs. 42% expected by chance; $z = 3.33$, $p < .01$) for parents' expectations regarding the first report card (Tables 5.37, 5.38).

As just noted, parents' expectations show significant matching with children's marks at the middle of first grade (the first report card) in all three areas. What is happening in those cases where parents' expectations and childrens' marks differ?

Most parents bet on a "B". Given the lack of variance in marks, it is hard for effects to display themselves (the power of tests is low) but nevertheless some interesting trends emerge.

The marks in reading at the middle of grade 1 were worse than parents expected for 17% of the children and better for 28%. The tendency for non-matching expectations to underestimate marks reaches significance at the .05 level ($X_1^2 = 4.19$). Parents would probably prefer to err on the side of expecting too little than too much. At the end of the year the marks in reading showed approximately the same marginals and bore almost exactly the same relations to parents' expectations as marks obtained earlier. Note that parents' expectations were only measured at about the middle of the year. Of those children who received A's in reading at the end of the year (N = 58), only 33% of parents had looked for a better performance. At this time, however, 33% of children (compared to 28% earlier) do better than their parents hoped and 18% (compared to 17% earlier) do worse than their parents hoped.

The marks in arithmetic at the middle of grade 1 were worse than parents expected in 20% and better than expected in 23%, about an equal division. At the end of first grade, marks were worse than parents expected in 12% and better in 34% of the cases (Table 5.40). The degree of asymmetry for the "unmatched" cases at the end of grade 1 is highly significant ($X_1^2 = 14.84$, $p < .01$). There is some improvement during the

Table 5.36

		Reading Mark, Time 1				
		1	2	3	Total	Percent
Parent Reading Expectation, Time 1	1	19	17	1	37	24.0
	2	31	60	8	99	64.3
	3	2	8	6	16	10.4
	4		2		2	1.3
	Total	52	87	15	154	
	Percent	33.8	56.5	9.7		100.0

Table 5.37

		Arithmetic Mark, Time 1				
		1	2	3	Total	Percent
Parent Arithmetic Expectation, Time 1	1	14	23	2	39	25.3
	2	17	67	6	90	58.4
	3	3	14	6	23	14.9
	4			2	2	1.3
	Total	34	104	16	154	
	Percent	22.1	67.5	10.4		100.0

Table 5.38

		Conduct Mark, Time 1					
		1	2	3	4	Total	Percent
Parent Conduct Expectation, Time 1	1	17	13	2		32	21.5
	2	29	55	13		97	65.1
	3	2	7	8	1	18	12.1
	4		1	1		2	1.3
	Total	48	76	24	1	149	
	Percent	32.2	51.0	16.1	0.7		100.0

year in arithmetic marks, and this probably accounts for the smaller percentage of children who do worse than parents expected. Thus the picture at the end of the year duplicates almost exactly that seen for reading, differing to some small extent because of differing marking practices.

In conduct at midyear 27% of the children do better than their parents expect and 19% do worse. At the end of the year 27% of the children again do better and 17% do worse (Table 5.41). In conduct children's performance is closer to what parents expect than is true for either substantive area. Parents may be cautious about pressuring children by expecting more than children can produce in reading or arithmetic. Their expectations may be more veridical in conduct where the child should have more direct control of his mark and where presumably direct parental pressure would be less damaging to performance.

Discrepances between Parents' Expectations and Mark at End of First Grade

Reading

Marks at the end of first grade in reading, in contrast to midyear marks, do not agree significantly with parents' expectations (Table 5.39). The percentage agreement is 49%, and chance expectancy is 43%. Significant agreement would require a percentage of 51% or more. Parents' expectations, unlike children's were sampled only once, sometime prior to the first report card. Parents' expectations for the end of the year, then, unlike children's, were not subject to modification by feedback during the year. (See the last portion of the preceeding section for further midyear-year end comparisons in all three areas.)

Arithmetic

Marks at the end of first grade in arithmetic (Table 5.40) agree very significantly with parents' initial expectations ($z = 2.67$, $p < .01$), continuing the significant agreement noted at midyear. The agreement, however, has dropped a few percentage points from 56% to 53%.

Conduct

Marks at the end of first grade in conduct agree better with parents initial expectations than marks at midyear (56% agreement vs. 54% agreement earlier). This year end agreement is highly significant ($z = 3.59$, $p < .01$) as was the earlier agreement.

In the three areas then, only for conduct do year end marks agree better with parents' midyear expectations than midyear marks. Both arithmetic and conduct marks remain significantly related to parents mid-year expectations while reading marks drop below significance.

Table 5.39

		Reading Mark, Time 2					
		1	2	3	4	Total	Percent
Parent Reading Expectation, Time 1	1	19	15			34	23.0
	2	36	50	10	1	97	65.5
	3	3	9	4		16	10.8
	4			1		1	0.7
	Total	58	74	15	1	148	
	Percent	39.2	50.0	10.1	0.7		100.0

Table 5.40

		Arithmetic Mark, Time 2					
		1	2	3	4	Total	Percent
Parent Arithmetic Expectation, Time 1	1	21	13	2		36	24.3
	2	33	52	2		87	58.8
	3	3	13	6	1	23	15.5
	4			2		2	1.4
	Total	57	78	12	1	148	
	Percent	38.5	52.7	8.1	0.7		100.0

Table 5.41

		Conduct Mark, Time 2					
		1	2	3	4	Total	Percent
Parent Conduct Expectation, Time 1	1	19	9	1		29	20.4
	2	26	55	10	2	93	65.5
	3	3	7	6	2	18	12.7
	4			2		2	1.4
	Total	48	71	19	4	142	
	Percent	33.8	50.0	13.4	2.8		100.0

How Parents' Expectations Compare in Different Areas

There is a high degree of concordance between parents' expectations in reading and arithmetic--63% agreement ($z = 4.43$, $p < .01$). This is about 18 percentage points above chance expectancy and so is significant in practical as well as statistical terms. (Children's concordance also exceeds chance significantly but by a smaller margin, 36% vs. 26%.) Interestingly, while children's expectations significantly favor reading over arithmetic, parents do not exhibit this bias. Of those parents whose guesses for arithmetic and reading do not agree, only about 53% favor reading over arithmetic.

The agreement between parents' guesses for conduct and arithmetic, like that for children's guesses, is not significant.

The agreement between parents' guesses for conduct and reading is not significant either. This finding differs from that for children whose guesses for reading tend to match guesses for conduct. Also parents again do not favor one area over another when their guesses disagree. (Children looked for reading to be higher than conduct when their expectations were not matched.)

How Does Parents' Expectation-Mark Agreement Vary with the First Mark Received?

Reading

Using the same paradigm as that used earlier with children's expectations one can investigate how parents' forecasting ability varies across levels of marks received. The marginal distribution for parents' expectations is A-24%, B-65%, C-10%, and D-1%. If no other force were at work, one would expect this set of guesses across each level of mark actually received. The distributions observed are given in Table 5.42.

Only in the case of B does the mode of any row fall in the "O" column. It looks as though <u>all</u> parents are playing safe--guessing a "B" no matter what the ability level of their children is. All entries in the "O" column do indicate some over representation. The smallest over representation (69% vs. 65%) occurs for the row based on the largest number of cases (87). The parents whose children get C's appear more accurate but the number of cases (15 in this row) is too small to warrant firm conclusions.

Arithmetic

A similar analysis can be applied to arithmetic. The marginal distribution for parents' expectations is A-25%, B-59%, C-15%, D-1%. As Table 5.43 shows, the mode for each row does not fall in the "O" column (as would happen if all parents accurately predicted marks),

Table 5.42
Parent Expectation-Mark Agreement by Level of
Mark Received in Reading

		Parent Expectation Minus Mark; Reading, Time 1								
		-3	-2	-1	0	+1	+2	+3	Total	Percent
Reading Mark, Time 1	1				37% 19	60% 31	4% 2		52	100%
	2			20% 17	69% 60	9% 8	2% 2		87	100%
	3		7% 1	53% 8	40% 6				15	100%
	4									
	Total		1	25	85	39	4		154	

Table 5.43
Parent Expectation-Mark Agreement by Level of
Mark Received in Arithmetic

		Parent Expectation Minus Mark; Arithmetic, Time 1								
		-3	-2	-1	0	+1	+2	+3	Total	Percent
Arithmetic Mark, Time 1	1				41% 14	50% 17	9% 3		34	100%
	2			22% 23	64% 67	14% 14			104	100%
	3		12% 2	38% 6	38% 6	12% 2			16	100%
	4									
	Total		2	29	87	33	3		154	

Table 5.44
Parent Expectation-Mark Agreement by Level of
Mark Received in Conduct

		Parent Expectation Minus Mark; Conduct, Time 1								
		-3	-2	-1	0	+1	+2	+3	Total	Percent
Conduct Mark, Time 1	1				35% 17	60% 29	5% 2		48	100%
	2			17% 13	72% 55	9% 7	2% 1		76	100%
	3		8% 2	54% 13	33% 8	5% 1			24	100%
	4			100% 1						100%
	Total		2	27	80	37	3		149	

but there is some movement in this direction for all ability levels. Again, while the pattern of over-representation in the "0" column is consistent, the smallest over-representation (64% vs. 59%) appears in the row with the most cases (104). The third row where the most interesting cases appear cannot shed much light because of the small number of cases (16). Parents seem a little less optimistic about arithmetic.

Conduct

For conduct marks the marginal distribution of parents' expectations is A-22%, B-65%, C-12%, D-1%. The agreement between parents' expectations and marks received is broken down by level of mark in Table 5.44. Although for the reasons outlined earlier one might expect the closest agreement for conduct, as can be seen in Table 5.44, there is only a slight departure from what would be expected under the assumption that the child's actual performance level has no effect. The magnitude of the over representation of the column headed "0" closely approximates the over representation that occurs for arithmetic (about 10% on the average). The case bases, however, are more uniform than for arithmetic.

To sum up: While parents do significantly better at forecasting children's marks than would be expected by chance the overall level of exact matching of forecasts and marks remains moderate. Exact matches occur in 55%, 57%, and 54% of the cases for reading, arithmetic, and conduct respectively. Is it parents of high, moderate or low ability students who are providing this extra accuracy in forecasting? In all three areas (reading, arithmetic, and conduct) parents exactly forecast children's marks more often than would be expected by chance, for children of all ability levels. In all three areas the smallest increment above chance levels occurs for the moderate ability students. Larger increments above chance occur for both the high and low ability students. The last two observations, while consistent across areas, may not attain statistical significance due to the generally smaller N's involved. For all levels of child ability the most frequently forecasted mark remains a "B" or "2".

How Does Parents' Midyear Expectation-Mark Agreement Vary with the Mark Received at the End of First Grade?

Tables relating the midyear discrepance between parents' expectations and marks to year end marks are Tables 5.45, 5.46, and 5.47. If a simple causal relationship existed between parent midyear expectations and year end marks one would expect an over representation on the minor diagonal of these tables. While a slight over representation is present in the cell representing a year end mark of "2" and the "Some" midyear child performance as parents expected for all these tables, the most notable cells are those in the upper left corners of the tables. The size of these upper left cells is anomalous. Why should children who

Table 5.45

		Reading Mark, Time 2			
		1	2	3	
Child Did B,S,W Than Parent Expected at Midyear	Better	24	14	3	41
	Same	26	45	10	81
	Worse	8	14	2	24
		58	73	15	146

Table 5.46

		Arithmetic Mark, Time 2			
		1	2	3	
Child Did B,S,W Than Parent Expected at Midyear	Better	17	12	4	33
	Same	29	50	5	84
	Worse	11	15	3	29
		57	77	12	146

Table 5.47

		Conduct Mark, Time 2			
		1	2	3	
Child Did B,S,W Than Parent Expected at Midyear	Better	20	11	4	35
	Same	22	44	10	76
	Worse	6	16	5	27
		48	71	19	138

did better than their parents expected them to do at midyear get the highest mark possible at year end? The lower parental expectations should produce some lower mark. A χ^2 test shows that only in reading and conduct are children who received a "1" at year end significantly more likely to have had parents who expected less--rather than more--in terms of their performance at midyear. (The respective χ_1^2 values are 7.03, $p < .01$; 0.55, N.S.; and 6.05, $p < .05$.).

The over representation can be understood in light of two other facts: First, if a child received a "1" at midyear he was very likely to also receive a "1" at year end. This occurs for 72%, 85% and 82% of the cases for reading, arithmetic and conduct respectively. And, second, parents' midyear expectations matched the children's high performance only about one half of the time. At midyear, of the children who got 1's the percentage of those whose parents expected 1's are 51%, 36% and 53% respectively for reading, arithmetic and conduct. Almost all the remainder of these parents had expected 2's.

These two conditions are enough to insure an over representation of the cells in question. This high stability of 1's insures that the roughly even split between parents expecting the same or a better mark (with few expecting a worse mark) displays itself at year end as well as at midyear. The stability of 1's itself might be taken as a sign of the causal ineffectiveness of parental expectations for those who received 1's but such an argument would have to assume parents' expectations behaved quite differently than children's expectations in response to feedback. Remember, children's expectations rose sharply following positive feedback but did not drop nearly as much following negative feedback. If parents' expectations behaved in a similar manner, parents of children who received a "1" would come to expect a "1", removing any "downward force" of parental expectations on those who received 1's at midyear. Unfortunately data to test the response of parental expectations to midyear mark feedback is not presently available.

How Does Parents' Expectation-Mark Agreement Vary with Parent Expectation Level?

As noted earlier there is a significant (above chance) tendency for parents' midyear expectations to exactly match their child's midyear mark. Is this above chance matching evenly distributed among parents independently of the level of performance they expect from their child?

Reading

The pattern of cases we would have expected for reading, given chance matching between parents expectations and child's midyear marks is presented in the following table:

Hypothetical Chance Distribution of Parent Expectation by Parent Expectation-Mark Agreement, Time 1

		Parent Expectation Minus Mark; Reading, Time 1							
		-3	-2	-1	0	+1	+2	+3	
Parent Expectation, Reading, Time 1	1		10	56	34				100%
	2			10	56	34			100%
	3				10	56	34		100%
	4					10	56	34	100%

This is arrieved at by placing the marginal distribution of children's marks (10%-3's, 56%-2's, and 34%-1's) in the appropriate section of each row. Given that parents show above chance levels of exact matching we would expect an over representation of the column headed "0" discrepance. The question here, however, is whether this over representation is stronger for those parents expecting any particular mark. Table 5.48 indicates the expected over representation of the column headed "0" occurs at all expectation levels, but this over representation seems to be stronger for those expecting either above or below average marks. Those parents expecting 1's exceed chance expectation by 17 percentage points and those parents expecting 3's exceed chance by 28 percentage points while those expecting 2's--an average mark--exceed chance by only 5 percentage points. (The row containing parents expecting 4's is omitted due to insufficient cases.)

Arithmetic

In arithmetic the marginal mark distribution is 10%-3's, 68%-2's and 22%-1's. In comparing Table 5.49 with a hypothetical table as constructed for reading--except using the arithmetic marginal percentages--we see that again the zero discrepance column is over represented at all levels of parental expectation but that this over representation is largest for those expecting both above and below average marks. Those parents expecting 1's exceed chance levels by 14 percentage points; those expecting 3's exceed chance by 16 percentage points; while those expecting 2's exceed chance levels by only 6 percentage points. (4's are omitted due to insufficient cases.)

Conduct

In conduct the mark distribution is 1%-4's, 16%-3's, 51%-2's, and 32%-1's. Table 5.50 indicates the hypothetical "0" discrepance column values are again exceeded at all levels of parents' expectations and that those parents holding above or below average expectations show the greatest excess. The excess for those expecting 1's is 21 percentage points, 3's is 28 percentage points and 2's is only 6 percentage points. (4's are omitted due to insufficient cases.)

Table 5.48
Parent Expectation-Mark Agreement by Level of
Mark Expected by Parents in Reading

		Parent Expectation Minus Mark; Reading, Time 1								
		-3	-2	-1	0	+1	+2	+3	Total	Percent
Parent Expectation, Reading, Time 1	1		3% 1	46% 17	51% 19				37	100%
	2			8% 8	61% 60	31% 31			99	100%
	3				38% 6	50% 8	13% 2		16	100%
	4						100% 2		2	100%
	Total		1	25	85	39	4		154	

Table 5.49
Parent Expectation-Mark Agreement by Level of
Mark Expected by Parents in Arithmetic

		Parent Expectation Minus Mark; Arithmetic, Time 1								
		-3	-2	-1	0	+1	+2	+3	Total	Percent
Parent Expectation, Arithmetic, Time 1	1		5% 2	59% 23	36% 14				39	100%
	2			7% 6	74% 67	19% 17			90	100%
	3				26% 6	61% 14	13% 3		23	100%
	4					100% 2			2	100%
	Total		2	29	87	33	3		154	

Table 5.50
Parent Expectation-Mark Agreement by Level of
Mark Expected by Parents in Conduct

		Parent Expectation Minus Mark; Conduct, Time 1								
		-3	-2	-1	0	+1	+2	+3	Total	Percent
Parent Expectation, Conduct, Time 1	1		6% 2	41% 13	53% 17				32	100%
	2			13% 13	57% 55	30% 29			97	100%
	3			6% 1	44% 8	39% 7	11% 2		18	100%
	4					50% 1	50% 1		2	100%
	Total		2	27	80	37	3		149	

Summary

In all three areas, reading, arithmetic and conduct, there is some above-chance matching between parents' expectations and their children's midyear mark, at all levels of parental expectation. Those parents, however, who hold either above or below average expectations for their children exceed chance to a greater degree than those parents who only expect an average performance. This observation may mean either a) that the degree of causal efficacy of parental expectation differs by expectation level (which, for example, might be made plausible by arguments concerning "salience of parental expectations for the children") or b) that parents with either above or below average children are more likely to have come to expect this above or below average performance from their children than parents of average children are to have come to expect an average performance from their children.

While no significance tests for the differences in excesses of exact matching at the different parental expectation levels were done, the consistent pattern indicates these observations are worthy of attention.

The Effects of Parents' Expectations on Changes in Children over the First Grade Year

Effects on Marks

Reading

Changes in reading marks over the first-grade year as a function of parents' expectations vs. first marks are summarized in Table 5.51. As earlier noted, the sum of the elements on the minor diagonal gives a measure of the extent to which children's marks move in a direction consistent with parents' expectations. The movement is highly significant ($z = 2.52$, $p < .01$). Looking at the four corners of the table one sees that the movement consistent with parents' expectations is impressive in both directions while that inconsistent is almost invisible (1 in each cell).

It is of some interest to compare Table 5.51 with the similar table (5.23) for children. The four corner cells there account for 50 cases, of which approximately two-thirds (36) move to increase agreement. In Table 5.51 the four corner cells account for 27 changes and of these 93% move consistently. What is perhaps most interesting is the size of the cell representing children whose marks move down to increase agreement, that is, those who earlier did better than their parents expected.

Arithmetic

Changes in arithmetic marks over the first-grade year as a function of parents' expectations are summarized in Table 5.52. Here also the

Table 5.51

		Change in Reading Mark, T1 to T2			
		Up	Same	Down	
Child Previously Did B,S,W Than Parent Expected in Reading	Better	1	28	12	41
	Same	18	50	13	81
	Worse	13	11	1	25
		32	89	26	147

Table 5.52

		Change in Arithmetic Mark, T1 to T2			
		Up	Same	Down	
Child Previously Did B,S,W Than Parent Expected in Arithmetic	Better	2	26	6	34
	Same	19	63	2	84
	Worse	14	14	1	29
		35	103	9	147

Table 5.53

		Change in Conduct Mark, T1 to T2			
		Up	Same	Down	
Child Previously Did B,S,W Than Parent Expected in Conduct	Better	1	24	11	36
	Same	12	58	7	77
	Worse	15	13	1	29
		28	95	19	142

movement toward agreement, tested by examining the sum of the elements on the minor diagonal, is highly significant ($z = 3.21$, $p < .01$). If the parent expected the child to perform more poorly than was the case at midyear, there is a trend for that expectation to be realized over the year--18% of those children have marks that move down. These cases account for two thirds of the instances where marks do move down over the year.

Conduct

Table 5.53 summarizes changes in conduct marks over the first-grade year in terms of whether first marks received were consistent with parents' expectations. Again the movement toward consistency is highly significant ($z = 4.50$, $p < .01$) and a large proportion of those whose marks move down (58%)are recruited from the group who performed better than their parents earlier expected.

A comparison with the similar table based on children's expectations for themselves (Table 5.29) again suggests that the press to make performance coincide with expectations seems slightly stronger for parents' expectations than for children's expectations.

In sum, children's year end marks in all three areas show significant trends toward agreeing with parents' midyear expectations. The over representation on the minor diagonal is consistent with a causal hypothesis viewing parents' expectations as the independent variable.

Effects on Expectations

Reading, Arithmetic, Conduct

There is some tendency over the first grade year for children's expectations to rise a little in all three performance areas. This change in children's expectations does not appear to be related to the child's midyear performance level evaluated against parents' expectations. If a parent expected less than his child delivered, in all three areas (see Tables 5.54, 5.55, and 5.56) there may be a slight tendency for the child's expectations to improve. The top left corner cell in all three tables is larger than the other three corner cells, but in no case is this association strong enough to attain significance. This seemingly counterintuitive observation might have arisen as follows: For the children who did better than their parents expected it is quite reasonable to expect a large proportion of these children to have done better than they themselves expected. That is, a large majority of parents expect a B (or "2"), therefore if the child's performance exceeded this the child probably received an A (or "1") at midyear, which in turn would suggest a fairly high likelihood that the child's performance would exceed his own expectations. If the child expected anything other than the top mark possible this would be the case. As noted in an earlier

Table 5.54

Child's Reading Expectation Change, T1 to T2

		Up	Same	Down	
Child Previously Did B,S,W Than Parent Expected in Reading (T1)	Better	11	24	6	41
	Same	24	40	16	80
	Worse	7	10	8	25
		42	74	30	146

Table 5.55

Child's Arithmetic Expectation Change, T1 to T2

		Up	Same	Down	
Child Previously Did B,S,W Than Parent Expected in Arithmetic (T1)	Better	10	17	6	33
	Same	25	38	21	84
	Worse	4	17	8	29
		39	72	35	146

Table 5.56

Child's Conduct Expectation Change, T1 to T2

		Up	Same	Down	
Child Previously Did B,S,W Than Parent Expected in Conduct (T1)	Better	14	17	7	38
	Same	31	28	16	75
	Worse	11	10	7	28
		56	55	30	141

section on feedback, if a child performed better than he expected, there was a strong tendency for his expectations to rise--which is what probably happens to the children being discussed here.

The "counterintuitive" observation has now been explained as follows: If a child did better than his parent expected his expectations may tend to rise because the child is likely to also have performed better than he himself expected--a condition previously shown to produce a rise in children's expectations. The cause of the expectation rise is then seen as the consequence of positive mark feedback for the child, and not something connected with the "relatively lower" parental expectation.

Sex Differences

Sex and Marks

The sex differences seen for individual cohorts are generally borne out more strongly in the combined data. More boys get the poorest grade (C) in reading in the middle of first grade than girls. The same is true for conduct. The odds are roughly 2:1 in both areas that a boy will get the lowest mark. The number of low marks actually issued in both areas is small, however, amounting to 11% of reading marks and 17% of conduct marks. In arithmetic low marks are about equally distributed between the two sexes at midyear.

By the end of first grade there is an even more marked tendency for boys to account for the poor marks in reading and conduct, while poor marks in arithmetic continue to be assigned equally to both sexes. The changes over the year by sex can be seen in Table 5.57.

At midyear the tendency for boys to receive the poorest marks in reading is not strong enough to be significant. By the end of the year, however, 81% of the poorest marks awarded in reading go to boys and this is a highly significant sex-related discrepance ($z = 2.62$, $p < .01$). The same pattern exists for conduct. By the end of the year, 85% of the poorest marks awarded go to boys, again a highly significant disparity by sex ($z = 3.46$, $p < .01$).

Turning now to the other extreme of the distributions of marks for reading, one can examine the awarding of A's: At midyear 32% of the boys get A's and 37% of the girls get A's, not significantly different ($N = 96$ for boys, $N = 82$ for girls). But over the year the picture changes. Ten more A's are awarded at the end of first grade than at midyear, and 9 of the 10 go to girls. Even with such a small number, there is a significant sex related difference ($z = 2.21$, $p < .05$). About the same number of C's are awarded on both occasions but, as already shown, boys receive a significantly larger number of C's than girls at the end of the year.

Table 5.57
Students Getting C or Less by Sex

	Reading		Arithmetic		Conduct	
	Midyear	Year End	Midyear	Year End	Midyear	Year End
Boys	14	17	11	8	20	23
Girls	6	4	11	7	11	4
Total	20	21	22	15	31	27

The conclusion is apparent--there is very little difference in marginal distribution of marks in reading at the end of the year compared to midyear and almost all of it is accounted for by the awarding of a few more A's. In all cases but one these A's are awarded to girls.

For conduct 7 more A's are awarded at the end of the year than were awarded in the middle. Here, however, boys are the recipients, with 6 of the 7 "new" A's going to boys. So in spite of the increased propensity for the poorest marks in conduct to go to boys, there is also a push for boys' marks in conduct to move up.

In arithmetic despite stability at the low end of the distribution of marks, there is relatively large improvement in the average mark given--32 more A's are awarded at the end of the year than at midyear. Nineteen of these "new" A's go to boys, and 13 go to girls, an association with sex that is not significant. There is a 53:47 split of A's in favor of boys at midyear and this split widens to 56:44 at the end of the year, but the division is not significantly different from a 50:50 split at either time. The trend, however, can be contrasted with that in reading where most of the improvement in marks was noted for girls.

Sex and Children's Expectations

What of the child himself in terms of his sex vs. his expectations over the first-grade year? At midyear, when expectations of children are sampled for the first time, females may be a little more cautious than males in both reading and arithmetic. Eight per cent of girls vs. 5% of boys expect the lowest mark (D) in reading (which none actually receives). Two per cent of boys and 8% of girls expect the lowest mark in arithmetic. By the end of the first year only 5% expect C's or D's in reading, divided about equally between boys and girls (compared to 19% and 17% respectively at midyear). More expect low marks in arithmetic--14% expect C's or D's and again the sex division is about equal.

But the decreases in expectations for A's over the year are sex-related. One more girl expects an A in reading at the end of the year and 8 fewer boys do. There is thus a very clear parallel between children's expectations and the teachers' marking behavior for, as was seen earlier, the increased number of A's awarded in reading go almost entirely to girls.

Just the opposite tendency may be seen in arithmetic expectations for A's by sex. Eight more boys expect an A at the end of the year than expected an A at midyear, and 3 fewer girls look for an A at the end. Again children's expectations are veridical with teachers' marking behavior for, as seen earlier, "new" A's in arithmetic are somewhat likelier to go to boys.

For conduct more boys--33% vs. 23% (N = 100, and N = 86 respectively)--than girls expect the two lowest grades at midyear. Also more girls than boys expect an A (52% vs. 41%). At the end of the year a large majority of girls (65%, N = 84) expect an A, whereas 43% of the boys (N = 100)

expect an A. This difference is significant ($\chi_1^2 = 9.26$, $p < .01$). Two more boys and 10 more girls look for an A at year end compared to mid-year. However, as seen earlier, of 7 "new" A's in conduct, 6 go to boys. It appears that girls' expectations are not veridical in the case of conduct.

Sex and Parent Expectations

Of those parents (19) who expect their children to do poorly in reading (C or D), 68% expect it of a son. The sex difference is not significant with a sample this size. Of those parents (27) who expect their children to do poorly in arithmetic, a majority (56%) expect it of girls. In conduct 21 parents expect their children to do poorly and 71% of them expect it of a son. This last trend is significant beyond the 10% level ($z = 1.75$). To sum up, although parents' expectations are in the culturally predicted direction for reading and arithmetic, trends here are not strong enough to be statistically significant.

The marginals for parents' expectations and the marginals for marks obtained in reading are quite similar in data for combined cohorts or in each cohort separately. Compared with the 12% of parents who look for poor marks in reading, 10% of all children actually get poor marks. But of the 12 boys whom parents expect to get low reading marks, only 4 actually do. Two boys whose parents expect them to get "4" get "2", and six boys whose parents expect them to get "3" get "1's" or "2's". On the other hand, 7 boys do get "3's" whose parents look for higher marks. There is about the same variance in prediction for girls, but only 4 girls get a mark of "3" (none gets "4").

In arithmetic 10 (12%) parents expect their sons to do poorly, and only 2 boys confirm these expectations. In arithmetic 15 (21%) parents expect their daughters to do poorly and 6 daughters confirm that. The numbers are small, but three tendencies exist: parents are more likely to expect girls than boys to do poorly in arithmetic; of those children whose parents look for poor performance, the large majority do well, but, if there is a tendency to conform with parents' expectations, it may be for girls. In conduct parents' expectations are more nearly fulfilled, for 6 of 14 boys whose parents expect them to get low marks actually do so. Of the boys who expect low marks in reading (20%, N = 83), about one-fourth have expectations consistent with their parents. A convenient summary of the picture for boys whose parents have low expectations is in the upper half of Table 5.58. What is noticeable are (1) the variability of children's expectations in the face of low parental hopes, and (2) the high percentage (67%) who get A's and B's for actual marks when parents expect low marks. Of boys whom parents expect to do poorly in reading, one-third (4 out of 12) get a C and 5 expect to do poorly.

The picture can also be viewed from another perspective. For children with low expectations, what are parents' expectations and actual marks? Of boys who themselves expect to do poorly in reading, relatively few (3 out of 18) actually do poorly but 5 have parents who expect them to do poorly. The data are summarized in the lower part of Table 5.58.

Table 5.58
Boys' Reading

Parents Expect	Child Expects*	Child Gets*
	A - 7	A - 2
C - 10	C - 2	B - 4
	D - 1	C - 4
D - 2	C - 1	B - 2
	D - 1	
12	12	12

* The letter grades and frequencies reported for child's expectation and marks are provided <u>within the specified categories of parent expectations</u>.

Child Expects	Parent Expects*	Child Gets*
	A - 3	A - 2
	B - 6	B - 9
C - 13	C - 2	C - 2
	D - 1	
	A - 1	A - 2
	B - 2	B - 2
D - 5	C - 1	C - 1
	D - 1	
18	17	18

* The letter grades and frequencies reported for the parent's expectation and child's marks are provided <u>within the specified categories of child expectations</u>.

Table 5.59
Girls' Arithmetic

Parents Expect	Child Expects*	Child Gets*
	A - 5	A - 1
C - 14	B - 7	B - 8
	C - 1	C - 5
D - 1	D - 1	C - 1
15	14	15

* The letter grades and frequencies reported for child's expectation and marks are provided within the specified categories of parent expectations.

Child Expects	Parents Expect*	Child Gets*
	A - 1	A - 2
C - 16	B - 10	B - 11
	C - 1	C - 3
	A - 1	
D - 6	B - 3	B - 4
	D - 1	C - 2
22	17	22

* The letter grades and frequencies reported for the parent's expectation and child's marks are provided within the specified categories of child expectations.

Again the numbers at this time are too small to justify conclusions but if a parent expects a child to do poorly the child's expectations agree in 42% of the cases, while if a child expects to do poorly the parent is less likely to agree (29%). In both cases a large majority of children actually do well--67% of children whose parents have low hopes get A's or B's, while 83% of children who themselves have low hopes get A's or B's.

Girls whose parents have low expectations for arithmetic do not themselves hold low expectations but 40% do get a low grade (almost the same kind of relation as for boys' reading). However, 60% get a good grade. If a girl expects to do poorly, her parents are not likely to agree, and only 23% actually get low grades. Again if a parent has low hopes it is more predictive than a child's having low hopes, but both are more often wrong than right.

This general pattern persists throughout the entire set of tables of this type for reading, arithmetic and conduct, as well as for both sexes.

CHAPTER 6
RESULTS FOR ONE FIRST-GRADE COHORT (1972-73)
LOWER-CLASS INTEGRATED SCHOOL (40% WHITE)

Tabular Summary

Means and standard deviations for variables used in subsequent analyses and cross-tabulations are given in Table 6.1. The mean IQ (Primary Mental Abilities) for 130 children for whom IQ data were available from school records (88%) is 104.65 with a standard deviation of 22.6. The data are given for the entire cohort and also separately for white students and black students. The overview will be given first for the entire cohort and then separately for its white and black portions.

Entire Cohort

(1) In the lower-class integrated school parents' expectations in all three areas are much lower than children's expectations (significant beyond the .01 level for reading, arithmetic, and conduct). The mean expectations of parents and children, respectively are 2.33 vs. 1.28 in reading, 2.23 vs. 1.74 in arithmetic, and 1.89 vs. 1.07 in conduct.

Expectations were obtained from only 74% of the parents, even though interviewers went repeatedly to the homes of lower-class children to contact the parents. In advance of the interviewer's call a note was sent home from school with children explaining that an interviewer would come to the home a day or two hence and also what the purpose of the interview was. As many as three call-backs were used to try to reach lower-class parents and care was taken to send a black interviewer to see black parents and a white interviewer to see white parents.

(2) Before their first report card children are exceedingly optimistic in reading and conduct. Their expectations in reading and conduct are significantly higher than their expectations in arithmetic ($p < .01$ for both paired t-tests).

(3) Children's average expectations in all three areas remain remarkably constant over the first-grade year. The largest difference, that for conduct, is only 0.16 of a grade point.

(4) Children's marks on the first report card are low. (The reading marks are particularly hard to evaluate at this time because teachers use different marking schemes). The arithmetic marks average 3.03 and conduct marks average 1.93.

Table 6.1

Means, Standard Deviations for First-Grade Cohort
Lower Class School, 60% Black

	Combined			Black			White		
	N	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.
Parents' Expectations--T1									
Reading	105	2.33	0.81	62	2.37	0.77	43	2.28	0.85
Arithmetic	106	2.23	0.76	62	2.15	0.74	44	2.34	0.78
Conduct	106	1.89	0.59	62	2.08	0.49	44	1.61	0.62
Child's Expectations--T1									
Reading	113	1.28	0.59	68	1.28	0.62	42	1.24	0.48
Arithmetic	113	1.74	0.62	68	1.65	0.48	42	1.83	0.70
Conduct	113	1.07	0.26	68	1.07	0.26	42	1.07	0.26
Child's Expectations--T2									
Reading	119	1.32	0.70	71	1.28	0.59	46	1.30	0.76
Arithmetic	120	1.75	0.81	71	1.75	0.82	47	1.79	0.81
Conduct	120	1.23	0.51	71	1.24	0.49	47	1.21	0.55
Child's Mark--T1									
Reading	72	3.15	0.85	43	3.16	0.84	29	3.14	0.88
Arithmetic	129	3.03	0.75	76	3.09	0.77	51	2.90	0.70
Conduct	129	1.93	0.62	76	2.03	0.61	51	1.77	0.59
Child's Mark--T2									
Reading	127	2.59	0.89	78	2.69	0.90	49	2.43	0.84
Arithmetic	127	2.77	0.97	78	2.89	0.99	49	2.59	0.91
Conduct	127	1.72	0.73	78	1.81	0.74	49	1.59	0.71

Table 6.1 (Continued)

Summary of Changes Over Year

	Combined			Black			White		
	N	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.
Mark Discrepance (T1 Mark minus T2 Mark)									
Reading	69	0.70	0.63	43	0.63	0.62	26	0.81	0.63
Arithmetic	121	0.22	0.74	74	0.18	0.73	47	0.30	0.75
Conduct	121	0.21	0.68	74	0.22	0.71	47	0.19	0.65
Expectation Discrepance (T1 Expectation minus T2 Expectation)									
Reading	95	-0.04	0.87	61	0.00	0.82	34	-0.12	0.98
Arithmetic	96	-0.09	0.93	61	-0.13	0.90	35	-0.03	0.99
Conduct	96	-0.18	0.56	61	-0.18	0.50	35	-0.17	0.66
Mark-Expectation Discrepance T1 (T1 Mark minus T1 Expectation)									
Reading	60	1.87	1.02	39	1.77	1.04	21	2.05	0.97
Arithmetic	108	1.32	0.83	66	1.41	0.86	40	1.15	0.70
Conduct	108	0.84	0.58	66	0.92	0.62	40	0.68	0.47
Mark-Expectation Discrepance T2 (T2 Mark minus T2 Expectation)									
Reading	114	1.24	1.03	69	1.38	0.99	45	1.02	1.08
Arithmetic	115	0.95	1.23	69	1.09	1.25	46	0.74	1.20
Conduct	115	0.46	0.81	69	0.54	0.83	46	0.35	0.77

Table 6.1 (Continued)

I.Q. Correlations

Average I.Q.	Combined N	Combined r	Black N	Black r	White N	White r
I.Q.[1] Correlations with Midyear Marks						
Reading	71	.411**	43	.500**	28	.302
Arithmetic	121	.476**	73	.525**	48	.388**
Conduct	121	.236*	73	.157	48	.275
I.Q. Correlations with Year End Marks						
Reading	124	.484**	76	.462**	48	.494**
Arithmetic	124	.427**	76	.401**	48	.436**
Conduct	124	.191*	76	.006	48	.406**

[1] I.Q. scores are Primary Mental Ability scores.

* $p < .05$

** $p < .01$

(5) Children's marks improve significantly in all three areas over the first-grade year ($p < .01$ for all three paired t-tests). Improvement is largest in reading (0.56), next largest in arithmetic (0.26), and least in conduct (0.21).

White (N = 57) and Black (N = 86) Children in Cohort

(1) White parents and black parents have comparable expectations for their first-grade children except in conduct where white parents look for a significantly better mark ($p < .01$). Expectations were obtained from 77% of white children's parents and from 72% of black children's parents. (The race of 5 children could not be ascertained.) Parents' expectations are uniformly lower than children's within each racial group.

(2) Before their first report card children of both races are highly optimistic about their forthcoming marks in reading and in conduct. Both races are noticeably less optimistic about marks to come in arithmetic, with whites on the average estimating an arithmetic mark 0.18 units lower than blacks (difference not significant).

(3) Children's average expectations over the year are remarkably constant no matter what the race of the child.

(4) Average reading marks are very close in the two races (3.14 vs. 3.16). Small differences (not significant) favor whites in arithmetic (2.90 vs. 3.09) and in conduct (1.77 vs. 2.03) (first report card).

(5) Children's marks over the first grade year improve more for whites than for blacks in both reading and arithmetic (for whites, 0.71 and 0.31 respectively, and for blacks 0.47 and 0.20 respectively). In conduct the changes are small (0.18 and 0.22) and close to one another.

Discrepances Between Initial Expectations and the First Report Card

First, a word is needed about how marks are assigned in general, and in particular for reading. The report card which children take home contains the explanation below: "Your child's progress is being measured in terms of his progress in reaching standards or levels that are considered appropriate for his age or years in school." Marking is thus done with respect to grade level performance, comparing the child's progress with average progress of children his age. It is important to note that the child's ability is not weighed in, so a child with low ability is being measured against the same performance scale as the child with high ability. (In the suburban school children were graded in terms of how their performance compared to their expected performance adjusted for ability level.)

Three different schemes were used for reporting marks in reading for this cohort: the number of Distar units completed, or the primer level (two series) of the book the child has completed. In the case of primer level, two different reading series were used. Since these series use the same letters, but the letters designate different levels of primers in the two series, the system is confusing. "P" in one system, for example, means a higher mark than "P" in the other.

To keep initial marks consistent with later marks, they were changed to a scale like that used for other subjects.

Reading

Close to half of the marks given (43%) in reading appear equivalent to a 4. Most of the children (73%), however, expect the highest mark in reading. Only 5% expect a C, and none expects lower. There is a marked contrast between the marginal distribution of children's expectations (strongly skewed toward the high end) and the marginal distribution of marks given by the teachers (strongly skewed toward the low end). This difference leads to a highly significant asymmetry for children's expectations to exceed the marks received ($\chi^2_1 = 54.0$, $p < .01$).

Since no marks equivalent to A's are given it may appear as if our recoding of letter marks to numerical marks is faulty and the whole numerical scale should have been shifted one position lower. This is not the case however. Both the teacher's comments on the students' report cards and the levels of student competence normally associated with the letter grades involved, suggest the teachers did indeed view the students as displaying a range of competence markedly skewed toward the low end. Such a rescaling would also have placed this set of marks completely out of line with those given at year end when the usual marking system was used. As it stands, not a single student has his performance exceed his expectation in reading and only 10% of the students were able to meet their expectations. (7% would have been expected to do so by chance alone.) A large fraction of those who receive the lowest mark (21 out of 26 or 81%) expect the highest.

Arithmetic

Most children expect a B (60%) in arithmetic. Expectations in general are high--over 95% expect the top two grades. Fewer students have hopes and marks that match than one would expect by chance alone (16.7% would be expected to match by chance but only 12% of the cases actually do match). The child's mark is apt to be less than he expects. In fact 29% of the children receive the lowest possible mark in arithmetic and 75% received the lower two marks. As with reading, in arithmetic there is a marked asymmetry ($\chi^2_1 = 91.04$, $p < .01$), the children expecting better marks than they receive. Of those whose mark fails to match, 99% of the mismatches are of the type where the mark is lower than the child hoped. Again one notes the large number of children (29%) who receive the lowest possible mark, compared to the small number expecting either of the two lowest marks (5%).

Table 6.2

		Reading Mark, Time 1					
		1	2	3	4	Total	Percent
Reading Expectation, Time 1	1		10	13	21	44	73.3
	2		5	5	3	13	21.7
	3			1	2	3	5.0
	4						
	Total		15	19	26	60*	
	Percent		25.0	31.7	43.3		100.0

* One class is missing reading marks.

Table 6.3

		Arithmetic Mark, Time 1					
		1	2	3	4	Total	Percent
Arithmetic Expectation, Time 1	1	1	16	12	9	38	35.2
	2		9	37	19	65	60.2
	3		1	1	1	3	2.8
	4				2	2	1.9
	Total	1	26	50	31	108	
	Percent	0.9	24.1	46.3	28.7		100.0

Table 6.4

		Conduct Mark, Time 1					
		1	2	3	4	Total	Percent
Conduct Expectation, Time 1	1	22	67	11		100	92.6
	2		6	2		8	7.4
	3						
	4						
	Total	22	73	13		108	
	Percent	20.4	67.6	12.0			100.0

Conduct

In conduct we find the only area in which teachers give out substantial numbers of the highest possible marks. Twenty per cent of the students are awarded 1's in conduct while 68% receive 2's and 12% receive 3's. Since 93% of the students expect 1's, however, the significant asymmetry between expectations and marks ($\chi^2_1 = 80.00$, $p < .01$) remains. The matching between marks and expectations (26%) is about what one would expect by chance (24%), given the distributions involved. There remains a strong tendency for children to receive a mark lower than the one they expect.

In summary the relationship between children's expectations and marks at the middle of grade one strongly reflects the pronounced positive skew in the expectation distribution. This produces an asymmetry in the relationship between marks and expectations in all three areas. The children expect a higher mark than they receive (or conversely receive a lower mark than they expect). This asymmetry is enhanced in the cases of reading and arithmetic by the fact that the mark distributions in these areas are negatively skewed.

In general students' expectations match the mark they receive with about the same frequency one would expect on the basis of chance. The matches slightly exceed those expected by chance in reading and conduct but the reverse is true for arithmetic.

Discrepances Between Expectations and Report Cards at the End of First Grade

Reading

The agreement between children's expectations for reading and the mark they receive at the end of first grade is slightly (but not significantly) below what would be expected by chance (11% vs. 14%). A relatively small proportion of the cases is involved since the children's reading expectations are highly skewed toward high expectations (81% expect a 1; mean = 1.32) while the mark distribution is more nearly symmetrical (mean = 2.59). Of the 89% of the cases where the student's expectation does not match his mark, a significant majority, 92% ($\chi^2_1 = 71.5$, $p < .001$), expect a mark higher than the one they receive.

Arithmetic

Agreement between marks in arithmetic at the end of grade one and expectations at that time is almost exactly at the chance level (22.3% expected by chance, 22.6% observed). As with reading, for arithmetic generally higher expectations prevail (mean = 1.75) than marks attain (mean = 2.77) though the expectation distribution is not as markedly skewed as in the case of reading. Forty-three per cent of the children still expect a 1 in arithmetic. As in reading the preponderance of high

Table 6.5

		Reading Expectation, Time 2					
		1	2	3	4	Total	Percent
Reading Mark, Time 2	1	7	3			10	8.8
	2	43	5	2	1	51	44.7
	3	32	2	1	2	37	32.5
	4	10	5	1		16	14.0
	Total	92	15	4	3	114	
	Percent	80.7	13.2	3.5	2.6		100.0

Table 6.6

		Arithmetic Expectation, Time 2					
		1	2	3	4	Total	Percent
Arithmetic Mark, Time 2	1	4	6	2		12	10.4
	2	16	15	4		35	30.4
	3	16	15	6	3	40	34.8
	4	13	10	4	1	28	24.3
	Total	49	46	16	4	115	
	Percent	42.6	40.0	13.9	3.5		100.0

Table 6.7

		Conduct Expectation, Time 2					
		1	2	3	4	Total	Percent
Conduct Mark, Time 2	1	44	8	1		53	46.1
	2	37	7			44	38.3
	3	11	5	1	1	18	15.7
	4						
	Total	92	20	2	1	115	
	Percent	80.0	17.4	1.7	0.9		100.0

expectations in arithmetic results in a significant majority of mismatched cases (83%) expecting a higher mark than they receive ($X_1^2 = 39.11$, $p < .001$).

Conduct

In conduct both the mark and expectation marginals are skewed toward the high end, with the expectation marginal being slightly more skewed than the mark marginal. As a result of this similarity in marginals a greater proportion of matches are expected (44%) than were expected for reading (14%) and arithmetic (22%). The observed matches (45%) do not significantly exceed this chance level of matching. Again asymmetry is present. Of those whose conduct marks and expectations differ at year end, a significant majority (84%) have higher expectations than marks ($X_1^2 = 29.34$, $p < .01$).

In summary, at year end most of the observed relations between marks and expectations result from the skew towards high expectations on the expectation marginals. This results in significant asymmetry between marks and expectations (children generally holding higher expectations than the marks they receive) in all three areas. It also results in a relatively small proportion of matches being expected by chance alone (except for conduct where a skewed mark marginal increases the proportion expected). The chance levels of matching are not exceeded in any of the three areas. The general year end pattern of: asymmetry due to skewed expectation marginals, small proportions of expected matches, and lack of significant deviation from expected matches, is the same as the general pattern observed between marks and expectations at midyear.

How Do Children's Expectations Change Over the First-Grade Year

Reading

The marginals for reading expectations sampled in the middle of the year and those sampled at the end are almost identical, with close to 80% of children expecting the highest mark possible on both occasions. There is no shift downward, for the same number move up as move down (15). Since 79% expect a "1" on both occasions, a high percentage of matches (62%) is expected by chance. The number of matches observed, 68%, is larger but not significantly larger. The lack of variance in both expectation distributions makes the question somewhat trivial.

Since about the same number of children get more optimistic as get more pessimistic, it appears that the marked discrepance between the child's initial expectation and the first mark he received has had little impact. The expectations of children remain very high--close to 80% expect

Table 6.8

		Reading Expectation, Time 2					
		1	2	3	4	Total	Percent
Reading Expectation, Time 1	1	61	9	3	2	75	78.9
	2	11	4		1	16	16.8
	3	2	1			3	3.2
	4	1				1	1.1
	Total	75	14	3	3	95	
	Percent	78.9	14.7	3.2	3.2		100.0

Table 6.9

		Arithmetic Expectation, Time 2					
		1	2	3	4	Total	Percent
Arithmetic Expectation, Time 1	1	17	13	5		35	36.5
	2	25	23	6	4	58	60.4
	3		3			3	3.1
	4						
	Total	42	39	11	4	96	
	Percent	43.7	40.6	11.5	4.2		100.0

Table 6.10

		Conduct Expectation, Time 2					
		1	2	3	4	Total	Percent
Conduct Expectation, Time 1	1	73	15	1	1	90	93.7
	2	4	1	1		6	6.2
	3						
	4						
	Total	77	16	2	1	96	
	Percent	80.2	16.7	2.1	1.0		100.0

the highest grade! The effect of feedback will be discussed in detail in subsequent sections.

Arithmetic

There is a modest increase in the variance of arithmetic expectations over the year. On first inquiry over 95% of the children expect a "1" or a "2"; on second inquiry about 84% expect those grades. On the other hand, 7 more children expect a "1" on second inquiry than did on first inquiry. The total number of persons shifting up exactly equals the total number shifting down (28 move in each direction) but 9 of those moving down register a decrease of 2 units.

The number of matches, 42%, is almost exactly what one would predict by chance. The most noteworthy fact is that a very small number of children look for a mark below a B on first inquiry, whereas 16% have lowered their hopes on second inquiry.

Conduct

There is some mild movement toward lower expectations for conduct grades, expectations moving from 1 to 2 over the year, but about 77% have exactly the same expectation later in the year as they had earlier. Again, as earlier, the picture in conduct is largely uninteresting because such a large percentage of the children expect A's.

To sum up: There is very little difference in the expectations at midyear and year end in either reading or arithmetic. A mild movement down in conduct is not interesting because overall expectations are still so high.

How Do Expectations Change over the Year in Relation to Initial Expectation Level?

Reading

About 79% had high expectations at midyear (1's) and almost the entire remainder expected 2's (17%). About equal numbers of children move down (Table 6.11) but of the 16 children with expectation 2 at midyear 73% (11) move up whereas of the 75 children with expectation 1 at midyear only 19% move down. The reader should note that movement from an expectation of a 1 can only be in a downward direction (i.e., to a 2, 3 or 4) while movement from a 2 can be either upward (to a 1) or downward (to a 3 or 4).

The preponderant trend then is that a very high percentage have extraordinarily high expectations to start with and tend to keep them, while those with not-so-high initial expectations are very likely to increase them ($p < .01$).

Table 6.11

Reading Expectations at Midyear

		1	2	Total
Expectations Move or Stay the Same	Stay Same	61	4	65
	Move	14	12	26
	Total	75	16	91

Table 6.12

Arithmetic Expectations at Midyear

		1	2	Total
Expectations Move or Stay the Same	Stay Same	17	23	40
	Move	18	35	53
	Total	35	58	93

Arithmetic

Fewer children have very high expectations in arithmetic at both times. Of those whose expectations change about as many move up as move down over the year. Again note that children with an initial expectation of a "2" can move up or down. As in reading, a greater proportion of children with 2's initially change their expectations (mostly upward) compared to those who initially had 1's but this difference is not significant (Table 6.12).

Conduct

In conduct expectations are so likely to be "1" to start with that it is not sensible to see how many stayed the same or changed as a function of initial expectations. Again, for those who had 1's the only possible movement is downward movement, which in this instance happened to be 23% of the cases.

How Marks Compare in Different Areas

First Report Card

Initial marks in reading and arithmetic show more consistency than would be expected by chance. Although 31% agreement would be predicted by chance, 53% is observed ($z = 4.04$, $p < .01$). Teachers then are rather consistent in the marks they assign in the two subjects. The average mark in arithmetic is 3.03 vs. 3.15 in reading at midyear, slightly but not significantly higher.

For reading and conduct the exptected agreement of the two marks is 24% while 31% agreement is observed. Though not significant this difference is in the direction suggesting a consistent assignment of marks between the areas.

The observed agreement between arithmetic marks and conduct at midyear is 22%, which is equivalent to that expected by chance (22%). There is no more consistency, then, between marks in arithmetic and conduct than would be expected by chance, and probably the same holds true for reading and conduct. It appears that teachers assess the behavior area separately from the substantive areas, yet assess the substantive areas similarly.

Marks at End of First Grade

At the end of the year the agreement between marks in arithmetic and reading has increased to 58%, and this exceeds chance (29%) by an even greater margin than at midyear ($z = 7.60$, $p < .001$). As noted elsewhere, teachers at the end of the year give both more high and low marks in

Table 6.13

		Arithmetic Mark, Time 1					
		1	2	3	4	Total	Percent
Reading* Mark, Time 1	1						
	2	1	13	7		21	29.2
	3		5	11	3	19	26.4
	4			18	14	32	44.4
	Total	1	18	36	17	72	
	Percent	1.4	25.0	50.0	23.6		100.0

*One class is missing Time 1 Reading Marks.

Table 6.14

		Arithmetic Mark, Time 2					
		1	2	3	4	Total	Percent
Reading Mark, Time 2	1	8	2	1		11	8.7
	2	5	26	20	2	53	41.7
	3		9	19	12	40	31.5
	4		1	1	21	23	18.1
	Total	13	38	41	35	127	
	Percent	10.2	29.9	32.3	27.6		100.0

Table 6.15

		Conduct Mark, Time 2				
		1	2	3	Total	Percent
Reading Mark, Time 2	1	8	2	1	11	8.7
	2	33	16	4	53	41.7
	3	11	21	8	40	31.5
	4	4	11	8	23	18.1
	Total	56	50	21	127	
	Percent	44.1	39.4	16.5		100.0

arithmetic compared to midyear (the standard deviation increases from .749 to .969) and there is a rise in average mark from 3.03 to 2.77. In reading the increase in marks is dramatic--from 3.15 to 2.59, with no change in variability. The agreement between marks in reading and conduct is what would be expected by chance alone, both the chance and observed matches being 25% of the cases (N = 127).

The agreement between arithmetic and conduct at the end of the year, 29%, exceeds expectation, 22%, and is significant at the .05 level ($z = 2.21$).

By the end of the year, then, there is a noticeable increase in homogeneity in marks for reading and arithmetic, and for the first time a significantly greater than chance matching between arithmetic and conduct. Reading marks remain related to conduct marks only at a level expected by chance.

How Do Marks Change Over the First-Grade Year?

Reading

The second reading mark shows 43% of the children falling in the lower two categories, getting 3's and 4's. About the same number (42%) were in the lowest category (4) for the first reading mark. The remainder at the second time point split about 1:3 between 1's and 2's respectively (Table 6.16). About 30% had received 2's earlier, none received 1's. The marking scale has thus been altered at both ends with fewer children at the end of the year getting the lowest possible grade and a modest number (15%) receiving the highest possible grade.

By the end of the year the majority of children (57%) receive the two highest marks because so many 2's are given (42%). There is a significantly greater than chance degree of consistency in marks over the year (39%; $z = 2.38$, $p < .05$). Despite this greater-than-chance matching of midyear and year end reading marks the majority of marks (61%) do not match at both times. All these mismatches have the year end mark exceeding the midyear mark which is not too surprising considering the pronounced skew toward lower marks that was present at midyear. (The asymmetry of matches is significant beyond the $p < .001$ level with $\chi^2_1 = 42.0$.)

Arithmetic

Teachers give higher marks in arithmetic as the year progresses. Whereas 26% of the children received 1's and 2's on first marking, about 40% received 1's and 2's later in the year (Table 6.17). About the same fraction of children (26% vs. 27%) get the lowest grade both times, and the bulk of these cases are provided by children who received the same mark on both occasions (19%). The overall improvement in arithmetic marks is significant--16 move down while 40 move up ($\chi^2_1 = 10.29$, $p < .01$).

Table 6.16

		Reading Mark, Time 2					
		1	2	3	4	Total	Percent
Reading Mark, Time 1	1						
	2	8	13			21	30.4
	3	2	12	5		19	27.5
	4		4	16	9	29	42.0
	Total	10	29	21	9	69	
	Percent	14.5	42.0	30.4	13.0		100.0

Table 6.17

		Arithmetic Mark, Time 2					
		1	2	3	4	Total	Percent
Arithmetic Mark, Time 1	1	1				1	
	2	9	15	6	1	31	25.6
	3	3	19	26	9	57	47.1
	4		1	8	23	32	26.4
	Total	13	35	40	33	121	
	Percent	10.7	28.9	33.1	27.3		100.0

The strongest tendency, nevertheless, is for marks to remain the same. For 54% of children marks remain the same. The degree of consistency is significantly greater than that expected by chance (30%) ($z = 7.22$, $p < .01$).

Conduct

Here as well there is considerable consistency in conduct marks over the year. About 50% of the children receive the same grade on both occasions with 37% expected to do so by chance (Table 6.18). There is, then, a significant consistency in conduct marks ($z = 3.10$, $p < .01$) from the middle to the end of the year. Of those children whose marks do shift, the shift is significantly in a positive direction (43 improve, 18 decline, $\chi^2 = 10.24$, $p < .01$) resulting in a general increase in marks over the year. Given that the marks were reasonably high at midyear, this results in the marks being skewed toward the high end at the end of the year.

In summary, then, higher marks are given at year end than at midyear in all areas. This results in a significant asymmetry in the tables presented. Or, equivalently, for the majority of the cases in which marks do not match at the two times, the mark received at year end is higher than that received at midyear. The general increase in marks reduces the skewness towards low marks in reading and arithmetic that was present at midyear and produces a skewness toward high marks in conduct at year end.

There is significantly greater than chance matching of marks at the two times in all three areas, despite the fact that approximately half the cases are mismatched on marks at the two times.

How Expectations Compare in Different Areas

At midyear there is significantly less agreement between children's expectations for reading and arithmetic (28%) than would be predicted by chance (37%) ($z = 2.29$, $p < .05$). This stems mainly from the different percentages of 1's and 2's expected in these two subjects (see margins of Table 6.19) for 78% of children expect a "1" in reading while 60% expect a "2" in arithmetic. The reader should note that there is significantly less than chance matching between expectations for reading and arithmetic, but significantly more than chance matching between marks in the two areas.

There is too little variance in children's expectations for conduct at the time of the first report card (93% expect 1's) confidently examine the agreement between conduct expectations and expectations in the substantive areas. At the end of the year the agreement between expectations for reading and for conduct, or between those for arithmetic and conduct, must be interpreted cautiously because of the lack of variance in conduct expectations. Eighty per cent expect "1's" in

Table 6.18

		Conduct Mark, Time 2				
		1	2	3	Total	Percent
Conduct Mark, Time 1	1	22	6		28	23.1
	2	33	31	12	76	62.8
	3		10	7	17	14.0
	Total	55	47	19	121	
	Percent	45.5	38.8	15.7		100.0

Table 6.19

		Child Arithmetic Expectation, Time 1					
		1	2	3	4	Total	Percent
Child Reading Expectation, Time 1	1	26	60	1	1	88	77.9
	2	10	5	3	1	19	16.8
	3	1	3	1		5	4.4
	4	1				1	0.9
	Total	38	68	5	2	113	
	Percent	33.6	60.2	4.4	1.8		100.0

conduct, and 18% expect "2's". In both cases the number of matches does not differ significantly from the number expected by chance.

What Is Impact of First Mark on Expectations Later?

Reading

Children's expectations at the end of the year are still exceedingly high--75% continue to look for the highest grade, and a very small percentage (10%) hope for the mark received earlier. This amount of agreement between expectations and earlier marks is well within the range of chance expectation (8%). Over 90% still look for 1's and 2's despite their teachers' having given no 1's and given only 30% 2's earlier. A very large proportion of those who received 3's and 4's still look for 1's and 2's (42 out of 47, close to 90%). The present analysis shows that feedback in the form of marks has little impact, for as already pointed out, expectations over the year are remarkably similar.

Arithmetic

Receiving a low mark in arithmetic earlier in the year (almost 75% received a 3 or 4) did not dampen expectations much at the end of the year. Just as was true for reading, the agreement of later expectations with earlier marks is at chance level (20% observed and 19% expected). A sizeable fraction of those who received 3's and 4's (over 41%) continued to hope for a "1" mark at the end of the year. In fact a greater proportion (43%) expected a "1" at year end than had expected a "1" at midyear (35%)! Relatively few children (about 7%) have expectations lower than the mark they received in arithmetic. More (about 73%) expect a higher grade than they received earlier. This degree of asymmetry around the main diagonal is highly significant ($\chi^2_1 = 60.84$, $p < .001$).

Conduct

Again, marks in conduct have a heavily skewed marginal with little variance to be explained. A sizeable fraction (80%) of children expect a "1" in conduct, a moderate reduction from the 93% who held this expectation earlier. It is interesting to note that this is the largest reduction in the percent of children expecting 1's in any of the three areas and that this is the only area in which the teachers' mark distribution was not skewed toward low marks! Matching of midyear mark with year end expectations (33%) is again not statistically different from the matches expected by chance (30%). Asymmetry around the main diagonal remains significant ($\chi^2_1 = 50.58$, $p < .001$).

In summary, the first marks the children received seem to have had little effect on their expectations. As noted previously the fact that the children's expectations are markedly skewed toward high expectations, while the teachers' mark distributions are not, means that the large

Table 6.20

		Reading Expectation, Time 2					
		1	2	3	4	Total	Percent
Reading Mark, Time 1	1						
	2	14	5	1		20	29.9
	3	16	1	1	1	19	28.4
	4	20	5	2	1	28	41.8
	Total	50	11	4	2	67	
	Percent	74.6	16.4	6.0	3.0		100.0

Table 6.21

		Arithmetic Expectation, Time 2					
		1	2	3	4	Total	Percent
Arithmetic Mark, Time 1	1	1				1	0.9
	2	13	12	5		30	26.8
	3	22	21	8	3	54	48.2
	4	12	11	3	1	27	24.1
	Total	48	44	16	4	112	
	Percent	42.9	39.3	14.3	3.6		100.0

Table 6.22

		Conduct Expectation, Time 2					
		1	2	3	4	Total	Percent
Conduct Mark, Time 1	1	21	4	1		26	23.2
	2	56	15	1	1	73	65.2
	3	13				13	11.6
	Total	90	19	2	1	112	
	Percent	80.4	17.0	1.8	0.9		100.0

majority of children receive negative feedback in the form of marks at midyear being lower than midyear expectations. A minority receive neutral feedback (mark equals expectation) and a rare few receive positive feedback (mark exceeds expectation). Expectations remain skewed toward high expectations despite the largely negative feedback.

The lack of any significant deviation from chance matching in any of the areas suggests that the midyear mark holds no special appeal for the children. There is no special (beyond chance) tendency for children to "adopt" their midyear mark and turn it into a year end expectation. A more detailed examination of feedback effects is presented later.

Does Children's Expectation-Mark Agreement Vary with Year End Mark Received?

The type of analysis carried out in the corresponding section of Chapter 5 cannot be carried out here. The combination of exceedingly high child expectations with relatively low mark assignments restricts the possible patterns of outcomes so severely that it is in general meaningless to attempt to assess whether children of different abilities differentially incorporate their actual performance (as viewed by the teacher) in the formation of their expectations. The extreme discrepances between children's expectations and marks suggests there is generally little incorporation of actual performance in the formulation of expectation, which leaves little hope of observing substantial numbers of reality-testing children at any particular ability level.

In the case most amenable to analysis--that of arithmetic at year end--there is no hint of reality testing for children of any ability level. The expected pattern based on a "marginal probability" model is as follows.

Hypothetical Pattern Based on a Marginal Probability Model
(Arithmetic at Year End)

		Expectation Minus Mark							
		-3	-2	-1	0	+1	+2	+3	
Mark	1				43	40	14	3	100
	2			43	40	14	3		100
	3		43	40	14	3			100
	4	43	40	14	3				100

The observed pattern approximates this very closely, indicating the children are not incorporating their performance (as viewed by the teacher) in the formulation of their expectations. The largest discrepance that occurs is for the high ability students (mark = 1), but this discrepance is in a direction opposite to that would be predicted on the basis of a reality testing model.

Table 6.23

Expectation Minus Mark, Arithmetic, Time 2

		-3	-2	-1	0	+1	+2	+3	Total	Percent
Arithmetic Mark, Time 2	1				33% 4	50% 6	17% 2		12	100%
	2			46% 16	43% 15	11% 4			35	100%
	3		40% 16	38% 15	15% 6	7% 3			40	100%
	4	46% 13	36% 10	14% 4	4% 1				28	100%
	Total	13	26	35	26	13	2		115	

Table 6.24

Expectation Minus Mark, Arithmetic, Time 2

		-3	-2	-1	0	+1	+2	+3	Total	Percent
Arithmetic Expectation, Time 2	1	27% 13	33% 16	33% 16	8% 4				49	100%
	2		22% 10	33% 15	33% 15	13% 6			46	100%
	3			25% 4	38% 6	25% 4	13% 2		16	100%
	4				25% 1	75% 3				100%
	Total	13	26	35	26	13	2		115	

Does Children's Expectation-Mark Agreement Vary with Expectation Level?

The question being asked here is: "Do children of differing expectations levels differ in their ability to form realistic expectations?" Are those who hold moderate expectations, for example, also those who form realistic expectations? On the surface it might appear that those who expect 1's are the most unrealistic given the low mark distribution. This is actually a misleading argument, however, since for tables like Table 6.24, realism would be indicated by a shift in the distribution of the discrepance between expectation and marks toward the "0" discrepance column. (This model is similar to that discussed in the preceeding section.)

The only case that is not too severely restricted (by the skewed mark and expectation marginals) to preclude analysis is again that of arithmetic at year end.

The marginal mark distribution for arithmetic is 1's = 10%, 2's = 30%, 3's = 35%, and 4's = 24%. If the mark distributions within each of the expectation levels are similar to one another and the marginal mark distribution (i.e., no expectation level displays realism) one would observe the following hypothetical pattern.

Table 6.24a
Hypothetical "Non-Realism" Pattern
Arithmetic Time 2

		Expectation Minus Mark Discrepance							
		-3	-2	-1	0	+1	+2	+3	
Expectation	1	24	35	30	10				100
	2		24	35	30	10			100
	3			24	35	30	10		100
	4				24	35	30	10	100

A shift toward the column headed "0" discrepance for any row would indicate the formation of realistic expectations for the level of expectations concerned.

The actual data, presented in table 6.24, indicate no such shifting. (Row 4 is not considered due to the small number of cases present.) The large majority of cells are within 3 percentage points of the expected values, the largest discrepance being 5 percentage points (row = 3, column = +1). The fit is surprisingly good considering the N's involved.

Our conclusion for arithmetic at year end must be that we have no evidence to suggest that any particupar expectation levels foster the formation of realistic expectations.

The Effects of Feedback on Marks

Reading

As noted earlier, there is an increase in the average reading mark from midyear to end of year (3.15 (N = 72) vs. 2.59 (N = 127)). In fact as Table 6.25 indicates, no child received a lower mark at the end of the year than he had at midyear. If, as could be the case, these initial reading marks are entirely uninterpretable by the child, one would expect no relation between improvement in marks and the midyear discrepance between expectation and mark. For the 57 cases where data are available this hypothesis is not contradicted as shown by a X^2 test on the collapsed table ($X^2_1 = 0.54$). The trend for the discrepant cases, however, in the non-collapsed table is suggestive. For 16 children with an earlier discrepance of 1, half improve (8) and half (8) stay the same; for 16 children with an earlier discrepance of 2, 63% improve; for 19 children with an earlier discrepance of 3, 74% improve. Though not statistically significant, a consistent trend such as this warrants attention.

Arithmetic

The change in marks over the first grade year can likewise be seen best in a collapsed table (Table 6.26). Again analysis is hampered because almost no one did better than he expected at midyear. Here however, of those--the large majority--who did worse than they expected, a large fraction (39%) get better marks and only 9% get lower marks. This is in contrast to what happens to children whose marks earlier were exactly what they expected. In this group (N = 12) 42% get lower marks and 17% show an increase. The numbers involved are small but an exact test used on the extreme cells shows that of those whose marks change up or down, change is significantly related to the direction of the midyear discrepance ($p < .01$, exact test) (Table 6.27). Almost all those whose marks improve, earlier got less than they hoped for, while of those whose marks equalled what they expected, a larger number got lower marks than higher at year's end.

Conduct

Over one-third of the conduct marks rise over the year. As was true in the other areas, no child received a mark higher than he expected at midyear. Looking only at children whose marks change, one sees that there is a significant tendency for those whose marks improve to be recruited from among students whose marks were less than they hoped for at midyear ($p < .056$, exact test).

Effects of Feedback on Expectations

Reading

If one looks closely at what happens to children's expectations at the end of the first grade year in reading in relation to the discrepancy earlier between expectation for the first mark and the first mark received,

Table 6.25 (Collapsed from Table 6-A-1)

		Expectation-Mark Discrepance Earlier		
		0-1	2+	
Year End Reading Mark is B,S	Better	12	24	36
	Same	10	11	21
		22	35	57

Table 6.26 (Collapsed from Table 6-A-2)

		Arithmetic Marks go			
		Up	Same	Down	
Child Earlier Did B,S,W Than He Expected	Better	0	1	0	1
	Same	2	5	5	12
	Worse	35	46	8	89
		37	52	13	102

Table 6.27 (Collapsed from Table 6.26)

		Arithmetic Marks go		
		Up	Down	
Child Earlier Did B,S,W Than He Expected	Better	2	5	7
	Worse	35	8	43
		37	13	50

Table 6.28 (Collapsed from Table 6-A-3)

		Conduct Marks go			
		Up	Same	Down	
Child Earlier Did S,W Than He Expected	Same	3	19	5	27
	Worse	34	31	10	75
		37	50	15	102

the following observations emerge. In 5 instances where the child's mark exactly equalled his expectations, 3 children's expectations later rose at the end of first grade and 2 children's expectations remained the same. In 50 instances where the child's expectations exceeded his mark, expectations remained stable in 31 instances, rose in 8 instances and declined in 11 instances. Thus in only 22% of the cases when the first mark was lower than the child hoped, does the child modify his hopes downward. Evidence disconfirming hopes does not lead to changing hopes. In no instance does the child's mark exceed his expectations! The changes in expectations for reading over the first grade year in terms of the midyear mark-expectation discrepance is best seen in a collapsed table (Table 6.29). Again the analysis is hindered by the lack of some midyear marks in reading. A rather small percentage (20%) registered an increase in expectations between midyear and end of year and of these 3 out of 11 cases had gotten what he expected at midyear. The others had gotter less than they expected at midyear. Almost everyone did worse than he expected, and all of those whose expectations go down are recruited from this group. Overall, however, the table is not very satisfactory because of the small number (5) of persons falling outside the "worse" row.

Arithmetic

The same kind of analysis for arithmetic expectations yields the following. There are 11 instances where earlier mark and earlier expectation exactly coincide. Of these, 4 children do not change their expectation and 5 hope for higher marks now. The remainder (2) are less optimistic than they were earlier. There are 84 instances where the earlier mark did not come up to the earlier expectation. Of these 36 (43%) maintain expectations at the same level and 22 (26%) increase. The remainder (31%) decrease. Only 18% however look for C's and D's although earlier the teachers have given 72% C's and D's.

Arithmetic expectations are more volatile than reading expectations--about 29% go up during the year compared to 20%, and 29% go down during the year compared to 20%. The greater upward movement is quite understandable because the midyear expectations for arithmetic were much more modest than reading expectations (1.74 vs. 1.28). The greater downward movement of the discrepant cases may be attributable to the children having a clearer understanding of their low marks in arithmetic compared to reading where the marking system was somewhat confusing at midyear.

Still, however, since only one child earlier did better than he expected, conclusions are limited. Looking only at the 55 children whose expectations changed (4 cells) (neglecting the single positive discrepance) one sees some indication for an earlier negative discrepance to lead to decreases in expectations. Numbers however are small and an exact test shows the relationship to be non-significant. The most likely event is for expectations to remain stable.

Conduct

Expectations in conduct tend to decrease slightly over the year (1.07 to 1.23). These expectations how more realism. Of 25 whose earlier

Table 6.29 (Collapsed from Table 6-A-4)

		Reading Expectations Go			
		Up	Same	Down	
Child Earlier Did B,S,W Than He Expected	Better	0	0	0	0
	Same	3	2	0	5
	Worse	8	31	11	50
		11	33	11	55

Table 6.30 (Collapsed from Table 6-A-5)

		Arithmetic Expectations Go			
		Up	Same	Down	
Child Earlier Did B,S,W Than He Expected	Better	1	0	0	1
	Same	5	4	2	11
	Worse	22	36	26	84
		28	40	28	96

Table 6.31 (Collapsed from Table 6-A-6)

		Conduct Expectations Go			
		Up	Same	Down	
Child Earlier Did S,W Than He Expected	Same	3	17	5	25
	Worse	1	57	13	71
		4	74	18	96

expectations matched the mark received, most (68%) expect the same mark again, 12% expect a higher grade, but 20% expect a lower mark. In 71 instnaces where the expectation exceeded the mark, most maintain the expectation (57). All of the rese, with one exception, lower their hopes. Again, however, in no case did the mark exceed the expectation. When expectation-changes are viewed in terms of the midyear discrepance between mark and expectation, the collapsed table (Table 6.31) summarizes the findings. For those whose expectations change, only 23% of the group, there is a tendency for expectations to be lowered when the previous mark was worse than expected.

In the cases of reading and arithmetic, expectations were more resistant to "bad news", for those changing moved about equally in the two directions. In conduct with 13 out of 14 decreasing, the trend is significantly downward ($z = 2.94$, $p < .01$) tested against a null hypothesis of changes up or down for those with negative discrepances.

Summary of Feedback Effects

For Marks

In all three sets of marks, those who improve tend to be those whose marks earlier were less than they expected. But those whose marks decline also tend to be those whose marks earlier were less than they expected simply because most of the children received marks lower than they expected at midyear.

However, in arithmetic and conduct significantly more students tended to show an improvement in their mark (as opposed to showing a decline) when there was an earlier discrepance between mark and expectations, compared to the improvements and declines that followed no discrepance. Too few cases with no discrepance are available in reading to place any confidence in the values observed there.

For Expectations

For the cases where expectations changed, in all three areas expectations tended to decline (non-significantly) more if their was an initial discrepance (over-estimation) than if there was no initial discrepance.

For the children whose marks and expectations do not coincide at midyear there is a tendency for the changes in marks and expectations that do occur to be changes in a manner that decreases the original discrepance. For the children with no mark-expectation discrepance at midyear, no consistent pattern of change emerges for either marks or expectations. Varying, but sizeable, proportions of cases retain the same mark or expectation at year end as at midyear no matter what the relationship between the mark and the child's expectation at midyear.

Do Marks and Expectations Show a Disproportionate Likelihood of Simultaneously Adjusting to Counteract a Mark Expectation Discrepance?

It was previously shown that both marks and expectations show slight tendencies to move over the course of the first year in a direction to reduce discrepances between marks and expectations at midyear. These observations could arise in several ways. First, a child could change both his mark and his expectation in a direction to reduce midyear discrepance. Second, a child could change either expectations or marks. That is, some children might maintain their initial expectations and effect a mark change, while other children might maintain their original mark and adjust their expectations.

Since all the discrepances reported in Table 6.32 are of the type where the child's expectation exceeds his mark, the column headed "Mark Up--Expectation Down" represents the first type of child described above. The second type of children--those who change either marks or expectations to reduce a discrepance--appear in the column headed "Mark Up--Expectation Same" (mark changes) and in the column headed "Expectation Down--Mark Same" (expectation changes). Comparing just these three columns (and neglecting all rows indicating no discrepance) one sees that results differ. Reading and conduct are rather similar in that the majority of the discrepance-reducing changes are of the mark-up-only type with the remainder of the cases about evenly split between the expectation-down-only and the mark-up-expectation-down types. The number of cases involved in these two patterns are 19, 4, and 1 for reading and 29, 4, and 4 for conduct. Arithmetic has approximately equal numbers of cases for all three types of patterns (13, 10, and 13 respectively).

In all three areas the simple change types (mark or expectation change) together are more numerous than the double change types (mark and expectation change). The safest general conclusion, given the small numbers involved and differences between areas, is that discrepance reduction occurs in various ways and data so far collected are not extensive enough to permit definitive findings.

As a final note, the reader should be cautioned that the preponderance of cases appearing in the columns headed "Mark Up" compared to the columns headed "Mark Down" does not necessarily imply the necessity of "adjusting for leniency of year end marking". If causal efficacy is granted to expectations over the course of the year, the skewed expectation distribution would also explain some (and possibly a large) part of the preponderance of increasing marks over declining marks.

Why does Feedback Have Little Effect on Expectations?

What may be happening to account for the "lack of reality testing" in reading? Low marks had almost no effect on children's expectations.

Perhaps the children do not know how they have been evaluated. The marks for reading as recorded on report cards are actually almost inscrutable. The authors were able to interpret these marks only after consultation with teachers and the principal. As mentioned earlier, three

Table 6.32
Mark-Expectation Discrepance at Time 1 by Mark and Expectation Changes Between Times 1 and 2: Reading, Arithmetic, and Conduct

Reading

		Mark Up			Mark Same			Mark Down		
		Expectations			Expectations			Expectations		
		Up	Same	Down	Up	Same	Down	Up	Same	Down
Mark Minus Expectation, Reading	3		10	3		2	3			
	2	1	6	3	1	5				
	1	4	3	1	2	5	1			
	0	2	1		1	1				

Arithmetic*

		Mark Up			Mark Same			Mark Down		
		Expectations			Expectations			Expectations		
		Up	Same	Down	Up	Same	Down	Up	Same	Down
Mark Minus Expectation, Arithmetic	3		1	1		4	2			
	2	2	5	7	2	7	3			
	1	5	7	5	9	11	5	4	1	3
	0		1	1	2	2		3	1	1

Conduct

		Mark Up			Mark Same			Mark Down		
		Expectations			Expectations			Expectations		
		Up	Same	Down	Up	Same	Down	Up	Same	Down
Mark Minus Expectation, Conduct	3									
	2		5			4				
	1	1	24	4		19	4		5	5
	0	2			1	14	3		3	2

* One case where a student's mark exceeded his expectations is omitted from this portion of the table.

separate schemes were used for marking which reflected progress in a given set or reading materials. None of the schemes coincides with the A, B, C, or 1, 2, 3 codes for other subjects on the report card. The Distar-units scheme reports the number of units completed. Only a person well versed in these instructional materials and in the general progress of first graders in reading could readily interpret this scale. The other two scales are literal--P, PP, etc.--to indicate the book the child is currently reading. Again, only a practiced reading teacher familiar with different genre of introductory reading texts could evaluate the child's progress. To complicate things further, the two series using P, PP, etc. use the same letters (e.g., P) to stand for two different levels of progress. A "Primer" level in one series is not equivalent to a "Primer" level in the other.

What children are receiving as feedback may therefore depend very little on recorded marks in reading. They may be relying almost exclusively on informal evaluations of the teacher given during classroom recitations. The Distar Program, interestingly enough, emphasizes lavish use of positive reinforcement: the teacher congratulating the child by shaking his hand, giving the child raisins or other small pieces of food, pinning signs on the child such as "I did well today", as well as the more usual types of reinforcement like a steady flow of verbal praise and smiles. This exaggerated kind of informal positive feedback may kindle the very high expectations the children continue to register at the end of the year and swamp any negative feedback from report cards, especially since the report cards are essentially uninterpretable to children and parents.

Encouragement and positive feedback are certainly necessary to motivate children in day-to-day classroom efforts, but one wonders what effect the extreme dissonance between the teacher's classroom behavior and her forced use of an absolute level grade scale may eventually have. What happens to the child, for example, when later in the second or third grade he realizes what his earlier report cards actually signified and looks back upon a long series of low marks?

While the uninteligibility of marking schemes is a plausible explanation for the lack of a feedback effect in reading, the lack of feedback effects in arithmetic and conduct are more readily understood, especially as the year progresses. Here a plausible explanation for the lack of effect may lie in the informal positive feedback given by the teacher in daily classroom sessions outweighing the infrequent negative feedback on report cards.

Children's Ability to Forecast

How Does Discrepance at Middle of Year Compare with Later Discrepance in Mark-Expectation Relations?

Reading

Table 6.33 displays how discrepances between expectations and marks behave over the first-grade year, or in other words whether children get

Table 6.33 (Collapsed from Table 6-A-7)

		Child Expected to do (Better, Same, Worse) Than He did in Reading			
		Midyear			
		B	S	W	
End of Year	B	38	2		40
	S	8	2		10
	W	4	1		5
		50	5		55

Table 6.34 (Collapsed from Table 6-A-8)

		Child Expected to do (Better, Same, Worse) Than He did in Arithmetic			
		Midyear			
		B	S	W	
End of Year	B	56	6		62
	S	18	3	1	22
	W	10	2		12
		84	11	1	96

Table 6.35 (Collapsed from Table 6-A-9)

		Child Expected to do (Better, Same, Worse) Than He did in Conduct			
		Midyear			
		B	S	W	
End of Year	B	38	4		42
	S	28	18		46
	W	5	3		8
		71	25		96

better at forecasting their marks. As before, no one at first marking gets marks higher than he expects--all get the same or lower. By the end of the year of those (50) who expected to do better than they did initially, 12--roughly one quarter--have a zero or positive discrepance, thus receiving a mark they expect or one better than they expect. The vast majority, however, who expected to do better earlier continue to find themselves receiving marks that do not live up to their expectations. There is some increase in the agreement between marks and expectations (9% got what they expected at midyear vs. 18% at year end), but the strong trend toward over-optimism is mitigated only moderately. There is some undershooting at year's end, about 9% expect to do worse than they actually do. These few are happily surprised.

Arithmetic

A similar analysis is presented for arithmetic in Table 6.34. This table has much the same pattern as the one for reading, but trends are clearer. First, there is considerable moderation in optimism by the end of the year--35% of children now show expectations the same or less than their mark compared to 13% at midyear. Second, a majority (58%) of the children repeatedly expected better marks than they received. And third, of those who change (from agreement to disagreement, or the reverse), 8 move away from agreement while 19 move toward agreement. A χ^2 test for related samples shows that this trend approaches the 5% level in significance ($\chi_1^2 = 3.70$). There is, in other words, a distinct movement that makes expectations agree with marks even though 65% remain over-optimistic.

It is of interest to note that the rate of movement away from midyear mark-expectation agreement (8 of 11 or 73%) is considerably larger than the rate of movement away from midyear disagreement (18 of 84 or 21%). That is, between midyear and year end students depart from the state of mark-expectation agreement at a faster rate than they depart from the state of mark-expectation disagreement. If these rates remained stable, more children would continually be found in a state of mark-expectation disagreement than in the state of agreement. There is, however, reason to believe that at least one of these rates will change. In particular if the first agreements occurred by chance (or factors unrelated to performance) the high proportion of followup disagreements would be expected. As agreements come to be a function of the child's actual perception of his performance, however, followup agreements should become much more likely. That is, the rate of departure from the agreement state should decline. The incorporation of evidence now being collected will indicate whether such a shift does occur during the second year and indeed will also provide a larger case base on which the rates for the first year can be calculated.

The reader should note that a similar pattern of a large migration away from agreement and a small migration away from disagreement occurs in the case of reading. The small case base for the agreement state precluded discussion.

Conduct

For conduct a summary is shown in Table 6.35. Again there is considerable movement toward agreement over the first-grade year with over-optimism decreasing from 74% to 44%. As before when the trend toward agreement is contrasted with the trend away from agreement, the trend toward agreement is confirmed, here at a high level of significance ($\chi^2_1 = 11.43$, $p < .001$).

In conduct, unlike in reading and arithmetic, the rate of movement away from the state of mark-expectation agreement is smaller than the rate of movement away from the state of mark-expectation disagreement (7 out of 25 or 28% vs. 28 out of 71 or 39%). In fact the rate of movement away from agreement is the smallest of the rates for the three areas and the rate of movement away from disagreement is the largest of the three areas. All these facts are consistent with an observation that children may be more able to obtain, comprehend, and incorporate information relevant to their conduct than to the two substantive areas. This interpretation is also suggested by the fact that conduct displays the lowest percentage of repeated over-estimations (40%) and the highest percentage of initial (midyear) agreements (26%) of the three areas.

In summary, there is evidence for arithmetic and conduct that children's accuracy in forecasting increases significantly. (Too few cases are available to draw a conclusion for reading.) The absolute amounts of accuracy, however, even at the end of the year are only about 30%, not overly impressive. Perhaps the most noticeable fact about all three areas is that at year end at least a few children forecast a lower mark than that received. No underestimates occur with respect to first marks. In conduct children display the least amount of repeated over-estimation (which none the less is substantial in all areas) as well as the greatest ability to maintain agreement between marks and expectations.

Parents' Expectations

Parents are optimistic but guarded. As noted in Table 6.1 parents' expectations are fairly stable across reading and arithmetic, 2.33 and 2.23 respectively. They are somewhat more optimistic about conduct, 1.89. They are also noticeably less variable for expectations in conduct, the standard deviation 0.35 being about half what is seen for reading or arithmetic. Most (49%) look for a B in reading, 52% look for a B in arithmetic and 69% look for a B in conduct.

Parents' Expectations vs. Children's Expectations

As already noted in several places the children's expectations for reading and conduct are extremely high. Because of this the congruence expected between parents and children in low (19%). The level of congruence actually observed (25%) is also low, not in fact significantly greater than would be predicted by chance.

Table 6.36

		Child's Reading Expectation, Time 1					
		1	2	3	4	Total	Percent
Parent's Reading Expectation, Time 1	1	9	1	1		11	12.4
	2	32	11		1	44	49.4
	3	22	3	2		27	30.3
	4	7				7	7.9
	Total	70	15	3	1	89	
	Percent	78.7	16.9	3.4	1.1		100.0

Table 6.37

		Child's Arithmetic Expectation, Time 1				
		1	2	3	Total	Percent
Parent's Arithmetic Expectation, Time 1	1	7	7		14	15.6
	2	20	26	1	47	52.2
	3	5	21	1	27	30.0
	4		2		2	2.2
	Total	32	56	2	90	
	Percent	35.6	62.2	2.2		100.0

Table 6.38

		Child's Conduct Expectation, Time 1			
		1	2	Total	Percent
Parent's Conduct Expectation, Time 1	1	19		19	21.1
	2	56	6	62	68.9
	3	7	1	8	8.9
	4	1		1	1.1
	Total	83	7	90	
	Percent	92.2	7.8		100.0

The agreement between parents and children is higher for arithmetic--36%--but this again does not differ from chance prediction of agreement (39%). The agreement between parents and children for conduct is 28%, despite the fact that 92% of the children look for a 1. This matching again does not significantly depart from the chance prediction of agreement (25%).

These data contradict the hypothesis that parents' expectations for school performance determine what school preformance the children come to expect for themselves at least very early in their school career.

Parents' Expectations vs. Marks

Reading

As noted earlier there is little correspondence between marks children expect in reading and those they receive. Likewise there is little correspondence between expectations of parents and the marks their children receive on their first report card for only 27% of the parents' expectations are borneout (compared to 25% expected by chance). Both this low level of expected matching and the asymmetry (parents' expectations generally exceeding marks) are a result of the low level of reading marks initially assigned. A rather surprising thing is that 3 parents (5%) expect their child to get the lowest mark and in all 3 instances this expectation is confirmed.

Later at the end of the first grade year the amount of agreement between parents' expectations and marks obtained inreading goes up to 52%. This of course is highly significant ($z = 4.42$, $p < .01$). It is somewhat surprising that parents' midyear expectations significantly match year end reading marks but do not significantly match midyear reading marks. The resolution of this anomaly is the fact that excessively low marks were granted in reading at midyear while a mark distribution only moderately lower than usual appears at year end.

At year end significantly more ($\chi^2_1 = 4.50$, $p < .05$) of those parents whose expectations are non-veridical are disappointed (67%) rather than the reverse. A surprisingly large percentage (7%) expect their child to do poorly and over half of these children do (4 out of 7).

Arithmetic

In arithmetic one finds at midyear only about chance agreement between parents' forecasts of marks and the marks their children obtain (29% match compared to 28% matches expected by chance). Again the mark distribution is quite low, 75% of the children receive 3's or 4's, resulting in parents' seemingly overestimating their children's midyear marks ($\chi^2_1 = 46.1$, $p < .01$).

At year's end the amount of agreement (32%) does not significantly exceed chance level (27%). Since the mark distribution only rises slightly by year end (60% now receive 3's or 4's) there remains a marked asymmetry in the mismatched parents' expectations ($\chi^2_1 = 18.25$, $p < .01$). More

Table 6.39

		Reading Mark, Time 1					
		1	2	3	4	Total	Percent
Parent Reading Expectation, Time 1	1		7	3	2	12	18.7
	2		12	11	10	33	51.6
	3		2	2	12	16	25.0
	4				3	3	4.7
	Total		21	16	27	64	
	Percent		32.8	25.0	42.2		100.0

Table 6.40

		Reading Mark, Time 2					
		1	2	3	4	Total	Percent
Parent Reading Expectation, Time 1	1	5	7	1	1	14	13.5
	2	5	30	11	4	50	48.1
	3		9	15	9	33	31.7
	4		1	2	4	7	6.7
	Total	10	47	29	18	104	
	Percent	9.6	45.2	27.9	17.3		100.0

Table 6.41

		Arithmetic Mark, Time 1					
		1	2	3	4	Total	Percent
Parent Arithmetic Expectation, Time 1	1		8	5	4	17	16.3
	2	1	11	28	11	51	49.0
	3		6	17	10	33	31.7
	4				3	3	2.9
	Total	1	25	50	28	104	
	Percent	1.0	24.0	48.1	26.9		100.0

Table 6.42

		Arithmetic Mark, Time 2					
		1	2	3	4	Total	Percent
Parent Arithmetic Expectation, Time 1	1	3	6	7	1	17	16.2
	2	8	17	15	12	52	49.5
	3	2	6	12	13	33	31.4
	4			1	2	3	2.9
	Total	13	29	35	28	105	
	Percent	12.4	27.6	33.3	26.7		100.0

Table 6.43

		Conduct Mark, Time 1				
		1	2	3	Total	Percent
Parent Conduct Expectation, Time 1	1	9	15		24	23.1
	2	13	50	8	71	68.3
	3		5	3	8	7.7
	4	1			1	1.0
	Total	23	70	11	104	
	Percent	22.1	67.3	10.6		100.0

Table 6.44

		Conduct Mark, Time 2				
		1	2	3	Total	Percent
Parent Conduct Expectation, Time 1	1	17	6	1	24	22.9
	2	31	28	11	70	66.7
	3	1	4	5	10	9.5
	4	1			1	1.0
	Total	50	38	17	105	
	Percent	47.6	36.2	16.2		100.0

children (51%) do worse than expected if the parents' forecast is incorrect, and only 17% do better. Parents are looking for more B's in arithmetic than teachers award (50% vs. 28%) and they look for very few failures but a large number are recorded (34% vs. 60%).

Conduct

Parents' expectations for conduct match the marks given in terms of marginal distributions--parents expect the teachers to behave as they do in assigninf marks here. There is in addition a better-than-chance ability of a parent to forecast his child's conduct mark, with 60% of the predictions being accurate. This accuracy exceeds chance (52%) at the 5% level of significance ($z = 1.97$). Also there is roughly an even division among parents who are over-optimistic and over-pessimistic in this area.

At the end of first grade ther is again significant matching between parents' expectations and children's marks in conduct (37% matching expected and 48% is observed; $z = 2.62$, $p < .01$). Here, however, despite this agreement parents' display under-optimism. Significantly more (35%) children exceed their parents' hopes than disappoint them (17%) ($X_1^2 = 5.89$, $p < .05$).

In sum while parents expectations consistently display above chance matching with marks in conduct and consistently display only chance levels of matching with marks in arithmetic, reading displays significantly above chance matching only at year end. All the cases of non-significant matchings are accompanied by discrepant mark and expectation marginal distributions--the discrepances being quite radical in the case of midyear reading and arithmetic. The marginal discrepances (low marks vs. moderate expectations) result in parents expecting higher marks than their children receive in reading and arithmetic at both midyear and year end. In conduct no symmetry is present at midyear while conservative asymmetry (lower expectations than marks) appears at year end.

Sex Differences

Midyear Marks

At first marking the assignment of marks in reading does not show association with sex. In arithmetic there is likewise no significant association between marks and sex, although a few more of the high marks go to girls and a few more of the low marks go to boys. In conduct there is a strong association ($X_1^2 = 7.24$, $p < .01$) between high marks and femaleness when the split between the sexes is compared for A's vs. C's only. Over two thirds of all the A's go to girls and more than two-thirds of the C's go to boys. While the extreme marks display a highly significant association with sex the overall mark distribution is only moderately significant ($X_2^2 = 7.25$, $p < .05$) due to the similarity of the bulk of the cases which receive an average mark.

Table 6.45

Conduct Mark, Time 1

		1	2	3	Total	Percent
Sex	Male	9	39	14	62	48.1
	Female	20	41	6	67	51.9
	Total	29	80	20	129	
	Percent	22.5	62.0	15.5		100.0

Midyear Expectations

In terms of children's expectations in reading, a few boys (about 10%) expect to do poorly, whereas only one girl (2%) does--otherwise the distributions are very similar. In arithmetic and conduct the expectations for the two sexes look almost identical.

Midyear Marks by Expectations

The agreement between expectations for reading and marks in reading is the same for the two sexes--agreement levels are almost identical with chance expectancy in both sexes. Also the tendency to over-optimism is strongly present in both sexes and present to just about the same degree.

In arithmetic expectations, over-optimism appears but is less pronounced than for reading. There is, furthermore, no tendency toward significant matching between expectations and marks in arithmetic for boys or girls. The tables for the sexes are very similar.

In conduct, as already noted, more boys than girls get the lowest grade, but the variance in expectations is so minimal that it is not possible for sex differences to appear.

In sum, boys and girls are virtually identical in the relationships displayed between midyear marks and expectations for all three areas.

Year End Marks

At second marking the assignment of marks in reading shows a tendency toward an association of higher marks with females. While this association does not attain conventional significance levels ($\chi^2_3 = 3.25$, N.S.) it is interesting that the assignment of the very highest and very lowest marks is almost exactly equivalent by sex of child, while the intermediate marks, 2's and 3's, are the areas in which the association is tending to appear.

At year end the assignment of marks in arithmetic favors girls (average girl's mark is 2.72 compared to boy's average of 2.83), but the difference is very slight.

A noticeable and large difference by sex occurs for marks in conduct. The highly significant association ($\chi^2_2 = 18.17$, $p < .01$) between sex and conduct marks in the expected direction--girls disproportionately receiving the high marks, boys the low. Girls receive two thirds of the high marks and boys receive 86% of the low marks.

Year End Expectations

In terms of reading expectations at the end of grade one, the distributions for the two sexes are virtually identical. For arithmetic expectations, however, in contrast to midyear when the two sexes were the same, there is now a tendency for girls to look for the higher grades and

Table 6.46

Reading Mark, Time 2

		1	2	3	4	Total	Percent
Sex	Male	6	22	24	11	63	49.6
	Female	5	31	16	12	64	50.4
	Total	11	53	40	23	127	
	Percent	8.7	41.7	31.5	18.1		100.0

Table 6.47

Arithmetic Mark, Time 2

		1	2	3	4	Total	Percent
Sex	Male	7	17	19	20	63	49.6
	Female	6	21	22	15	64	50.4
	Total	13	38	41	35	127	
	Percent	10.2	29.9	32.3	27.6		100.0

Table 6.48

Conduct Mark, Time 2

		1	2	3	Total	Percent
Sex	Male	18	27	18	63	49.6
	Female	38	23	3	64	50.4
	Total	56	50	21	127	
	Percent	44.1	39.4	16.5		100.0

boys to look for the lower. This association is nonsignificant ($\chi^2_3 = 5.19$, N.S.), however. No sex differences in conduct expectations appear at the end of the year, and the midyear pattern is thus maintained.

In sum, for sex differences so far, the sexes seem almost identical in terms of both midyear and year end marks and expectations for reading and arithmetic. In conduct, while the expectations are the same, girls tend to get higher marks than boys at both midyear and year end.

At midyear boys and girls seem to display nearly identical relations between their marks and expectation--both being at about the chance level of matching and asymmetric in that expectations generally exceed marks.

Sex Differences in Parents' Expectations, and Congruence of Expectations Between Parents and Children

Reading

Parents expect boys and girls to do about the same in reading. They expect boys to average 2.37 and girls to average 2.30. The sex distributions do not differ much within mark categories, and the difference is nonsignificant.

In terms of how parents' expectations agree with children's expectations within sex groups, there is almost exactly chance agreement for boys but significantly better than chance agreement for girls ($z = 2.27$, $p < .05$). While girls' and parents' expectations coincide to a significantly greater than chance degree, and boys' and parents' expectations do not, the absolute level of matching remains low in both cases (27% for girls, 22% for boys). In all cases for girls when parent and child disagree, girls expect to do better. Indeed, there is more skewing of girls' expectations, for none expects less than a B, whereas 10% of boys expect a C or D. Boys, if they differ from their parents, are prone to have higher expectations for reading on their first report card ($\chi^2_1 = 19.53$, $p < .01$).

Arithmetic

For arithmetic parents' expectations are almost identical for the two sexes, with a slight shading (2.17 vs. 2.28) in favor of boys. The average expectations are close to those in reading for girls, but are slightly higher for boys.

In terms of how parents' and children's expectations agree, again for boys there is no significant concordance (39% observed and 41% expected). As with reading, in cases of disagreement between parent and child there is a much stronger tencency for a boy's expectation to exceed his parents' than the reverse ($\chi^2_1 = 5.76$, $p < .05$).

Table 6.49

Parent's Reading Expectation, Time 1

		1	2	3	4	Total	Percent
Sex	Male	9	20	18	5	52	49.5
	Female	5	30	15	3	53	50.5
	Total	14	50	33	8	105	
	Percent	13.3	47.6	31.4	7.6		100.0

Table 6.50

Parent's Arithmetic Expectation, Time 1

		1	2	3	4	Total	Percent
Sex	Male	9	26	16	1	52	49.1
	Female	8	26	17	3	54	50.9
	Total	17	52	33	4	106	
	Percent	16.0	49.1	31.1	3.8		100.0

Table 6.51

Parent's Conduct Expectation, Time 1

		1	2	3	4	Total	Percent
Sex	Male	11	35	6		52	49.1
	Female	13	36	4	1	54	50.9
	Total	24	71	10	1	106	
	Percent	22.6	67.0	9.4	0.9		100.0

For girls the concordance with parents for arithmetic marks is at chance levels (37% observed vs. 36% expected). There is thus no significant above chance agreement between parents and girls as there was for reading. The tendency for girls' hopes to exceed parents' is seen again here even more strongly --of the 31 unmatched cases 94% exceed parents' hopes (X_1^2 = 21.81, $p < .01$).

Conduct

Surprisingly parents expect about the same marks for boys as for girls in conduct (1.90 for boys and 1.87 for girls). Because children's expectations in conduct are so markedly skewed, it is pointless to look at concordance with parents or to test the significance of the marked asymmetry.

Sex Differences in Mark Received vs. Parents Expectations

As noted earlier there is not a significant degree of congruence between parents' expectations and the marks their children actually receive in reading. This lack of agreement is noted for both sexes. Parents' forecasts are at chance levels for boys and girls (26% correct for boys, 27% correct for girls).

For arithmetic parents' expectations and children's marks also agree no more than would be expected by chance and there is no differential accuracy of prediction by sex (24% observed vs. 25% expected for boys, and 35% observed vs. 30% expected for girls). Asymmetry appears in the tables comparing parents' reading and arithmetic expectations with marks (expectations generally exceeding marks) due to the low marks assigned in these areas at the middle of first grade.

In conduct a picture emerges indicating that parents' predictions for girls are more accurate than for boys. The agreement between parents' forecasts for boys' conduct marks and the marks received is at a level attributable to chance (52% observed vs. 53% expected). For girls, however, the agreement (67%) significantly exceeds chance (51%) with $z = 2.74$, $p < .01$.

To sum up the last two sections: Parents expect boys and girls to do the same in all three areas. Only for reading do parents' expectations and girls' expectations for themselves agree more often than would be expected by chance. Parents and boys' expectations do not coincide in any area. In terms of agreement with marks received, parents' forecasts are at chance level for both sexes in both reading and arithmetic. Parents can forecast girls' conduct marks with better than chance success, but not boys'!

Black/White Differences

About 60% of children in the first grade classes are black. Two teacher are white, two black.

Differences in First Marks

Reading

When marks are given for the first time, there is no association between the child's race and his mark in reading. The proportion receiving various marks is roughly matched to the racial division. In particular the numbers of blacks and whites receiving the lowest mark exactly matches the margins on race.

Arithmetic

Arithmetic marks show some tendency toward association with race. Half of the highest marks assigned (2's) go to whites (who compose 40% of the students). Over 70% of the 4's assigned go to blacks (who compose 59% of the students). There is about a 10% bias evident at the two ends of the marking scale, with a tendency for whites to get the higher marks and blacks to get the lower marks. A χ^2 test of association based only on the two extreme marks (N = 65) gives the value 2.99, significant only beyond the .10 level.

Conduct

The assignment of marks in conduct shows an even stronger racial bias of the same sort, for 55% of the 1's are given to whites and 78% of the lowest marks (3's) are given to blacks. (No 4's are given to anyone.) There is a 15-18% bias at the two ends of the scale. A χ^2 test performed only on the extreme cells gives a value of 4.74, which is significant beyond the 5% level.

Marks at End of Grade One

Reading

At the end of the first-grade year, marks in reading begin to show an association with race not present earlier. There are more 2's given to whites and more 3's and 4's given to blacks than the margins predict. The association between level of mark and race, however, does not differ significantly from chance ($\chi^2_3 = 6.27$, $p < .10$). Another way to see the type of association by race that may be developing here, is to note that the agreement between marks given on the two occasions for whites is at chance level (31%) but for the 69% manifesting disagreement, <u>all move up</u>. For blacks there is a similar upward movement for all the unmatched cases <u>but</u> the percentage agreement (43%) significantly exceeds chance (26%, $z = 2.39$, $p < .05$). Teachers are more likely to give a black child the same mark he received earlier or a higher mark but they are more likely to give a white child only a higher mark.

Table 6.52

Reading Mark, Time 1

		2	3	4	Total	Percent
Race	White	9	7	13	29	40.8
	Black	11	12	19	42	59.2
	Total	20	19	32	71	
	Percent	28.2	26.8	45.1		100.0

Table 6.53

Arithmetic Mark, Time 1

		2	3	4	Total	Percent
Race	White	15	26	10	51	40.8
	Black	15	34	25	74	59.2
	Total	30	60	35	125	
	Percent	24.0	48.0	28.0		100.0

Table 6.54

Conduct Mark, Time 1

		1	2	3	Total	Percent
Race	White	16	31	4	51	40.8
	Black	13	47	14	74	59.2
	Total	29	78	18	125	
	Percent	23.2	62.4	14.4		100.0

Table 6.55

		Reading Mark, Time 2					
		1	2	3	4	Total	Percent
Race	White	4	27	11	7	49	39.2
	Black	7	25	28	16	76	60.8
	Total	11	52	39	23	125	
	Percent	8.8	41.6	31.2	18.4		100.0

Table 6.56

		Arithmetic Mark, Time 2					
		1	2	3	4	Total	Percent
Race	White	5	19	16	9	49	39.2
	Black	7	19	25	25	76	60.8
	Total	12	38	41	34	125	
	Percent	9.6	30.4	32.8	27.2		100.0

Table 6.57

		Conduct Mark, Time 2				
		1	2	3	Total	Percent
Race	White	26	17	6	49	39.2
	Black	29	33	14	76	60.8
	Total	55	50	20	125	
	Percent	44.0	40.0	16.0		100.0

Arithmetic

The picture changes little from the picture at midyear, with again whites getting more 2's (about 10%) and blacks getting more 4's (about 10%) than would be predicted form the margins. The distributions of 1's and 3's however is almost exactly what the marginals would predict so an overall X^2 test shows no significant association ($\chi^2_3 = 4.17$, N.S.). There is a significant matching of later marks with earlier marks for both races (55%, $z = 3.49$, $p < .01$ for whites; 53%, $z = 3.90$, $p < .01$ for blacks), and also a significant tendency for marks to improve in both races. Thirty four per cent of whites get higher marks in arithmetic at the end of the year than in the middle ($\chi^2_1 = 4.76$, $p < .05$) and 33% of blacks do so ($\chi^2_1 = 4.97$, $p < .05$). The picture in terms of movement or change over the year in arithmetic looks very similar for whites and blacks.

Conduct

At the end of the year a few more 1's go to whites and 3's to blacks than one would predict, but a X^2 test performed on the extremes of the table is not significant ($\chi^2_1 = 1.14$, N.S.). Earlier it was significant. There is a tendency for marks in conduct of both blacks and whites, if the marks cahnge, to change upward but this reaches conventional significance levels for blacks only ($\chi^2_1 = 3.04$, N.S. for whites; $\chi^2_1 = 5.02$, $p < .05$ for blacks). There is more agreement (no change) for whites than would be predicted by chance (55% agreement vs. 39% predicted, $z = 2.21$, $p < .05$). There is not significantly more agreement for blacks (46% agreement vs. 38% predicted). To sum up for conduct, at midyear more of the lowest marks (compared to highest marks) went to blacks. At the end of the year both groups have moved up (significantly in the case of blacks) but relatively fewer blacks remained at the same level. About the same percentage of blacks get 3's on each occasion (19% initially, 18% at year end), but a few more whites get 3's at the end (12%) than in the middle (6%) of the year.

To sum up, marks in general show no association with race at either midyear or year end. In several places whites seem to be disproportionately assigned high marks and blacks low marks but this only attains significance for midyear conduct. The pattern of changes in marks over the first year (as indicated by asymmetry) are the same for both blacks and whites in all areas. All asymmetry is of the type where year end mark exceeds midyear mark. There are differences by race in the matching of marks at the two times. Only blacks show significant mark matching over the year in reading while only whites show significant mark matching in conduct.

Race and Children's Expectations Before the First Report Card

There is almost no difference by race in expectations for reading. Most of both races expect the highest mark, 80% of whites and 81% of blacks have this expectation. In arithmetic there is also little difference in

expectations by race although the few children (4%) who do have low expectations are all white. One-hundred per cent of the blacks look for an A or B in arithmetic. The two distributions of conduct expectations by race are also almost identical, with virtually all (92%) in both groups looking for an A. Of the 8% who look for less than an A, all expect B's.

In sum: black and white children do not differ in their expectations.

How Do Initial Expectations Conform with First Marks for Blacks and Whites?

Reading

For both racial groups the agreement between guesses of marks received is within the range of chance expectancy (5% for whites and 7% for blacks). For white children in 20 out of 21 cases, the mark received is lower than that expected. For black children 34 out of 38 children receive a lower mark than expected. The most noticeable discrepancy for both races is the large number expecting 1's who get 4's, 43% of whites and 32% of blacks.

Arithmetic

There is not much agreement between the arithmetic mark expected and the mark received for whites or blacks. The agreement for whites is 10% with 19% predicted, which is within the range of chance. For blacks it is 11% with 15% predicted--again within the range of chance. About 10% of whites guess they will receive a 3 or 4, but no black child looks for less than a 2 in arithmetic. In both groups most children are looking for a higher mark than they receive. Teachers gave whites better marks than blacks in arithmetic, 20% of whites vs. 33% of blacks receive 4's, but this difference is not statistically significant.

Conduct

In conduct the picture is uninteresting, again because there is no variance in children's expectations. Ninety-three per cent of whites and 92% of blacks look for the highest possible mark. Teachers use more high marks in this area than in reading or arithmetic, giving 95% of the whites 1's or 2's and 86% of the blacks 1's or 2's. The difference between assignment of high and low marks to blacks and whites is significant at the .05 level using Fisher's exact test.

In sum, in all areas for both races only a small amount of matching between marks and expectations is expected by chance due to the skewed marginals (marks being low, expectations high) and the observed matches do not differ significantly from these low expected values. The skewed marginals produce a strong asymmetry (expectations exceeding marks) in all areas for both races.

How Consistent Are Expectations Between the Middle and the End of First Grade for Blacks and Whites?

Reading

For children of both races the consistency between midyear and end of year reading expectations is about 65% (62% for whites and 73% for blacks). This consistency, however, is not significant since it is forced by the lack of variance in children's guesses on both occasions. (67% agreement for whites and 66% for blacks would be expected by chance.) There is about equal movement in both races to increase or decrease expectations from one time to the next. The probable lack of intelligibility to children or parents resulting from the complex marking system in reading, discussed earlier, is relevant also in accounting for the high consistency of expectations. No events have occurred to disrupt consistently high expectations.

Arithmetic

There is 40% agreement in expectations for arithmetic between the middle and end of first grade for white children. This agreement is exactly what chance predicts. There is 42% agreement for blacks, with 41% predicted by chance. Children's expectations in both groups move up and down about equally during the first-grade year, but the movements in black children's expectations are more noticeable because none of them expected less than a 2 at the first interview. At the second interview 17% expect a 3 or 4. About 9% of white children on first interview expect a 3 or 4 and 11% on second interview expect a 3 or 4. The numbers involved are too small to warrant testing.

Conduct

For both races, there is little variance, but both groups tend to modify their high hopes to some degree. At midyear 94% of whites expect a "1", and 83% persist in this hope at the end of first grade. At midyear 93% of blacks expect a "1", and 80% have this expectation at the end of the year. Matching is high but not significant.

In sum: blacks and whites behave almost identically in terms of expectations compared over the first year. Highly skewed marginals result in high chance agreements between expectations at midyear and year end. Observed levels of matching do not differ from those expected by chance.

Racial Differences in the Influence of Marks on Expectations

Reading

Both races remain overly optimistic about reading marks after receiving low marks at midyear. The percentage of children holding a year end expec-

tation that is two or more points above their midyear mark is 54% for whites and 68% for blacks. Since white children received slightly (not significantly) higher marks at midyear these percentages seem quite in line with a statement that whites and blacks are identical in their adjustment of expectations to earlier marks. (The magnitude of the mark discrepance between the races was such as to give whites about 10% more 2's (the top mark assigned) and 6% fewer 4's than blacks.)

Arithmetic

Results here again appear about the same for children of both races. Thirty-one per cent of whites and 48% of blacks hold a year end expectation 2 or more points above their midyear mark. Again, however, whites had received slightly (not significantly) better marks at midyear. They received 15% fewer 4's and 11% more 2's (the highest mark assigned) than did the blacks. This marking discrepance may not be enough to completely offset the observed over-optimism difference. If it is not--which is hard to judge due to the small N's involved--blacks would seem to exhibit more over-optimism (less realism) than whites.

Conduct

Teachers give both races mostly 2's in conduct (64% for whites, 66% for blacks) but blacks receive 15% 3's and 19% 1's compared to whites' 4% and 31% respectively. Both races look for 1's at year end (82% for whites, 80% for blacks). Blacks are over-optimistic at year end in 15% of the cases while whites are over-optimistic in only 4% of the cases. It seems the differences in assigned marks are sufficient to account for the differences in over-optimism between the races--but again the N is small and conclusions are only tentative.

To sum up: blacks consistently display a greater probability of over-optimism at year end (i.e., year end expectations exceed midyear mark by 2 or more points) but this difference is largely accounted for by the fact that blacks generally received slightly more low marks. (At year end about 80% of children expect 1's in reading and conduct, and 40% expect 1's in arithmetic, for both races.) The most likely conclusion is that the races do not differ in their response to first marks--but the N's involved are small and as more data become available the black children's disproportionate over-optimism may attain significance.

Matching, which in this case indicates the adopting of a midyear mark as a year end expectation, occurs at chance levels in all areas for both races.

Teacher Differences

Fortunately two of the first-grade teachers were white and two black. This permits some limited comparisons in terms of same-race and cross-race effects between teachers and students, and also allows some inter-teacher comparisons.

Differences in First Marks

Reading

One black teacher and one white teacher assign approximately equal numbers of 2's, 3's, and 4's. The second white teacher gave no marks in reading at the time of the first report card. The second black teacher marked only about half the class and those who received a mark tended to get the lowest mark.

Arithmetic

The marks given in arithmetic vary considerably among teachers. Two white teachers have distributions that are not widely different (5, 17, 8 and 9, 16, 10, for 2's, 3's, and 4's respectively). One black teacher marks harder than all the other teachers, and gave a "4" to slightly less than half the class. The other black teacher assigned the most 2's and 3's of all the teachers. Thus the two black teachers bracket the white teachers, one being noticeably easier and one noticeably harder than the white teachers. The difference in mark distributions (11, 15, 4, and 7, 12, 15) between the two black teachers is significant ($X_2^2 = 7.36$, $p < .05$).

Conduct

In conduct the black teacher who was an "easy" marker in arithmetic, used one category (2) almost exclusively. The other black teacher's distribution closely resembles that for the two white teachers. All give about the same proportion (15%) of 3's (the lowest mark used), a few more 1's (about 30%) and the majority 2's (about 55%).

In sum while a few teacher differences in marking practices appear (both in severity of marking and range of marks used) the data do not show strong differences in marking practices by race of teacher. There are as yet too few cases for firm conclusions, however.

Children's Expectations at Midyear

There is not much variance among classrooms in terms of children's reading expectation, for as noted in other analyses, a very large percentage (78%) expect the highest mark. There is slightly less optimism in the class of the black teacher who turns out to be most lenient in marking--63% of the children in her group expect the highest mark compared to an average of 83% in the other three rooms. It may be important for a teacher to convey "realistic" expectations to children, thus avoiding serious disappointment as this teacher does.

In arithmetic children's expectations are more modest, 34% looking for the highest mark. About 60% of children expect a "2". These percentages do not vary much across classrooms, or by race-of-teacher, as the table below indicates:

		Arithmetic Mark Expected, Time 1		
		1	2	3
Race of Teacher	Black	16	29	3
	White	18	31	4

All children held high expectations for conduct and there is little variation from one classroom to the next.

Children's Expectations at the End of First Grade

For three teachers the distributions of children's expectations in reading look similar on both first and second elicitation. The fourth teacher, who is white, has most of the children in the entire grade (5 of 7) who have low expectations (for a 3 or 4) at the end of the year. At midyear no children in her class expected a 3 or a 4. Interestingly in this class the regular teacher was absent a large fraction of the year. The number of children involved however, is too small to be more than suggestive.

Changes in the Number of Children Expressing Different Arithmetic Expectations from First to Second Elicitation

		Expectation			
Teacher	Race	1	2	3	4
1	(w)	5	-3	3	1
2	(w)	8	-10	1	0
3	(b)	0	-5	1	1
4	(b)	-2	-4	6	0

The table above summarizes movement in children's arithmetic expectations over the first grade. It shows mainly an increase in the number of children expecting 1's in the classrooms of white teachers, and a decrease in the number of children expecting 1's and 2's from the black teachers. As already noted, one black teacher (3) assigns more high marks than either white teacher and one (4) assigns fewer. The one who assigned lower marks accounts for almost one half of the downward movement in expectations.

There is some downward shifting in conduct expectations for all teachers with perhaps a slightly stronger trend for the white teachers. Again, however, a very high proportion (80%) expect the highest mark so there is little variance to analyze.

In sum while there appear to be a few differences in the distributions of expectations for pupils of the different teachers at both midyear and year end these do not seem to be strong differences associated with the race of teachers.

Teacher Differences in Children's Expectations and Marks by Child's Race

White teachers induce slightly (but not significantly; $X_2^2 = 5.45$, N.S.) higher initial expectations for reading in children of both races (1.14 for white children and 1.07 for black children) than black teachers (1.32 for white students and 1.41 for black students). In arithmetic and conduct teachers of both races appear to arouse expectations of about the same level in both races of children (white teachers for white children--1.86 in arithmetic and 1.10 in conduct; white teachers for black children--1.70 in arithmetic and 1.00 in conduct; black teachers for white children--1.74 in arithmetic and 1.05 in conduct; black teachers for black children--1.62 in arithmetic and 1.15 in conduct).

So few children possess low expectations for reading at the end of first grade year (only 6% of all children) that racial comparisons are not warranted. In arithmetic children with low expectations at the end of first grade are slightly more numerous (17%) but there is not a significant difference associated with race. In conduct all children but 3% expect 1's or 2's (mostly 1's) so no racial comparisons can be made.

Numbers are small when double-race comparisons on marks are attempted. One overall pattern of double-race comparisons, however, emerges despite a few minor deviations. The general trend is for no interaction between race of child and race of teacher to appear. Black and white teachers treat black students similarly. Black and white teachers treat white students similarly. At times black students are treated differently than white students (e.g., in year end reading blacks tend to get lower marks) but the lower marks for blacks are assigned by both black and white teachers.

In sum: comparison of marks and expectations at both midyear and year end within categories of both teacher race and child race shows no interaction effects. No special effects on either marks or expectations seem to result from children participating in classrooms where the teacher's race is the same as the child's race, or where the teacher's race is different than the child's race. Small N's are involved in the individual tables but the consistency of the pattern suggests the general conclusion is likely to hold as more cases are obtained.

Table 6-A-1

Change in Reading Grade From Time 1 to Time 2

		Increases				Decreases				
		3	2	1	same	1	2	3	Total	Percent
Expectation-Mark, Reading, Time 1	-3		4	10	5				19	33.3
	-2		1	9	6				16	28.1
	-1		1	7	8				16	28.1
	0			4	2				6	10.5
	+1									
	+2									
	+3									
	Total		6	30	21				57	
	Percent		10.5	52.6	36.8					100.0

Table 6-A-2

		Change in Arithmetic Mark From Time 1 to Time 2								
		Increases				Decreases				
		3	2	1	same	1	2	3	Total	Percent
Expectation-Mark, Arithmetic, Time 1	-3			2	6				8	7.8
	-2		3	11	15				29	28.4
	-1			19	25	7	1		52	51.0
	0			2	5	5			12	11.8
	+1				1				1	1.0
	+2									
	+3									
	Total		3	34	52	12	1		102	
	Percent		2.9	33.3	51.0	11.8	1.0			100.0

Table 6-A-3

		Change in Conduct Mark From Time 1 to Time 2								
		Increases				Decreases				
		3	2	1	same	1	2	3	Total	Percent
Expectation-Mark, Conduct, Time 1	-3									
	-2			5	5				10	9.8
	-1			29	26	10			65	63.7
	0			3	19	5			27	26.5
	+1									
	+2									
	+3									
	Total			37	50	15			102	
	Percent			36.3	49.0	14.7				100.0

Table 6-A-4

Change in Reading Expectation From Time 1 to Time 2

		Increases				Decreases				
		3	2	1	same	1	2	3	Total	Percent
Expectation-Mark, Reading, Time 1	-3				12	3	2	1	18	32.7
	-2			2	11	1	1	1	16	29.1
	-1			6	8	2			16	29.1
	0		1	2	2				5	9.1
	+1									
	+2									
	+3									
	Total		1	10	33	6	3	2	55	
	Percent		1.8	18.2	60.0	10.9	5.5	3.6		100.0

Table 6-A-5

Change in Arithmetic Expectation From Time 1 to Time 2

		Increases				Decreases				
		3	2	1	same	1	2	3	Total	Percent
Expectation-Mark, Arithmetic, Time 1	-3				5	2	1		8	8.3
	-2			4	12	7	3		26	27.1
	-1			18	19	8	5		50	52.1
	0			5	4	2			11	11.5
	+1			1					1	1.0
	+2									
	+3									
	Total			28	40	19	9		96	
	Percent			29.2	41.7	19.8	9.8			100.0

Table 6-A-6

Change in Conduct Expectation From Time 1 to Time 2

		Increases				Decreases				
		3	2	1	same	1	2	3	Total	Percent
Expectation-Mark, Conduct, Time 1	-3									
	-2				9				9	9.4
	-1			1	48	12		1	62	64.6
	0			3	17	4	1		25	26.0
	+1									
	+2									
	+3									
	Total			4	74	16	1	1	96	
	Percent			4.2	77.1	16.7	1.0	1.0		100.0

Table 6-A-7

		Expectation-Mark, Reading, Time 2								
		-3	-2	-1	0	+1	+2	+3	Total	Percent
Expectation-Mark, Reading, Time 1	-3	2	9	5	1	1			18	32.7
	-2	1	6	5	2	1	1		16	29.1
	-1		2	8	5	1			16	29.1
	0			2	2	1			5	9.1
	+1									
	+2									
	+3									
	Total	3	17	20	10	4	1		55	
	Percent	5.5	30.9	36.4	18.2	7.3	1.8			100.0

Table 6-A-8

		Expectation-Mark, Arithmetic, Time 2								
		-3	-2	-1	0	+1	+2	+3	Total	Percent
Expectation-Mark, Arithmetic, Time 1	-3	4	3		1				8	8.3
	-2	2	8	6	8	2			26	27.1
	-1	4	10	19	9	7	1		50	52.1
	0		3	3	3	1	1		11	11.5
	+1				1				1	1.0
	+2									
	+3									
	Total	10	24	28	22	10	2		96	
	Percent	10.4	25.0	29.2	22.9	10.4	2.1			100.0

Table 6-A-9

		Expectation-Mark, Conduct, Time 2								
		-3	-2	-1	0	+1	+2	+3	Total	Percent
Expectation-Mark, Conduct, Time 1	-3									
	-2		4	5					9	9.4
	-1		5	24	28	5			62	64.6
	0			4	18	2	1		25	26.0
	+1									
	+2									
	+3									
	Total		9	33	46	7	1		96	
	Percent		9.4	34.4	47.9	7.3	1.0			100.0

CHAPTER 7
DISCUSSION

The analyses and results presented in Chapters 3 through 6 are lengthy and are also at times fairly complex, as when changes in performance or expectations are studied as a function of several predictor variables or as a function of prior differences in predictor variables. There are, nevertheless, many additional analyses not reported here or even yet undertaken because of time limitations, analyses dealing with self-esteem, sociometric standing, absences from school, and others. The writers tried to deal first with analyses which seemed most crucial to the causal interpretation of variability in children's expectations and marks. The reader is warned that subsequent analyses may lead to some revisions in the preliminary findings presented here.

Data for the middle-class school aggregate two successive first-grade classes (Cohort 1 and Cohort 2) and so allow fairly extensive analyses. More detailed breakdowns by sex, however, cannot be carried out even with two cohorts aggregated and must await addition of more cases by a future cohort.

Data for the lower-class school are more restricted, being based on a single cohort of first graders. Since all black-white comparisons stem from this cohort, the reader should bear in mind that the findings with respect to racial differences are both limited and tentative.

So far all performance data for children are based on teacher's marks in reading, arithmetic and conduct. Attempts to assess "Language" and "Spelling" with Cohort 1 in the middle-class school proved unworkable because the children were unable to understand what these terms meant. These topics were therefore dropped. In second grade and later, standardized test scores will be studied as well as the influence of expectation variables on standardized achievement measures.

School Differences

Several factors differ between the two schools besides social class. These factors are therefore confounded with class differences. To give some notion of population differences in the two communities served by the middle- and lower-class schools, data from the 1970 U. S. Census for census tracts served by the schools are given in Table 7.1. The lower-class school serves most of four tracts, the middle-class school most of three. In terms of median of head of household education, the social class composition of the two schools represents the customary definitions of middle class (some college) and lower class (high school or less). Income in the middle-class commonly is over twice that of the lower-class community.

Table 7.1

	Middle-Class School				Lower-Class School				
Tract Number	1	2	3	Ave.	1	2	3	4	Ave.
Median Years School Completed	12.4	13.6	12.9	13.0	10.2	8.4	10.2	8.5	9.4
Mean Income	$13,294	$24,101	$17,036	$18,143	$ 8,573	$ 6,487	$ 8,074	$ 8,165	$ 7,825
Median Value Owner-Occupied Houses	$19,200	$30,600	$28,200	$26,000	$ 6,300	$ 6,000	$10,200	$ 5,800	$ 7,075
Per Cent Black	0.0%	0.0%	0.2%	0.1%	63.7%	92.4%	7.3%	8.1%	18.46%
Per Cent of All Families with Income Below Poverty Level	3.2%	1.1%	2.0%	2.1%	20.2%	36.2%	15.8%	14.2%	21.6%

The middle-class school is white segregated, its regular staff is entirely white, its classes have no aides. The school building was recently enlarged and renovated. It has a large library, a cafeteria where hot food is prepared on site at noon, and a large playground. It is located in a quiet residential neighborhood of well-landscaped individual homes worth $30,000-$50,000 at current prices.

The lower-class school has about 60% black children, an integrated (roughly 50-50) staff, an aide in all kindergarten and first grade classes. The school is set on a busy street corner in a densely populated urban residential area where traffic is very heavy. Nearby are "row houses" worth $6,000 to $20,000 depending on exact location and internal condition, and many have been turned into apartments for 3 or more families. Parts of the school building were erected before the turn of the century. The school has its own playground with much play equipment and also a large gymnasium within the building and a large library. Hot lunches are delivered daily to the school. Many children in this school have family incomes low enough to qualify for free lunches under a federal program as Table 7.1 suggests. The school has the services of a social worker two days a week, a full-time guidance counselor, and "resource teachers" who help classroom teachers with reading instruction.

The two schools are very close to the same size with 3 to 4 first-grade classes in each depending on total enrollment. Staff quality in both schools is high and in both places there are many teachers who have taught at the school for several years. If anything, formal credentials of the staff in the lower-class school may be superior. Class size is about the same in the two schools.

Both schools have kindergartens with half-day sessions attended by many of the children who later enter first grade. Very little study of these kindergartens has been made so far, but one difference between schools very important for this research, is that the lower-class school gives report cards to kindergarteners. We attempted to query kindergarteners about their expectations for report cards but were unable to procure responses we felt were valid. In many cases the child would not respond at all; in other cases it was clear the child did not understand the nature of the task. Thus the meaning of "first report card" may not be exactly the same from one school to the other.

The reader must be cautious in attributing differences to social class or to intergration effects when there are so many other factors which also differ between schools. Fortunately with some notion of inter-cohort differences available for the middle-class school, one has a benchmark to gauge the size of intra-school variability.

I.Q. Differences

There is a significant yet relatively small difference between average I.Q.'s (PMA) of blacks and whites in the lower-class school (4.8 points--106.2 for whites and 101.4 for blacks). The I.Q. differences

between schools is also small, 115 for the middle-class school and 103 in the lower-class school. These two averages are significantly different but in second grade at the middle-class school (the only second grade for which there are data) the average I.Q. for Cohort 1 turns out to be 104 on the Stanford Test of Mental Maturity. School-to-school differences, then are no larger than year-to-year differences between repeated tests of (approximately) the same children in a single school.

I.Q. is a poor predictor of school performance in the early grades in the middle-class school. In first grade the highest correlation there is around .20 at midyear. A correlation of this magnitude implies that only about 4% of the variance in performance is explained by I.Q. differences. Correlations at the end of the year between I.Q. and reading, arithmetic, or conduct diminish and are non-significant in every case.

The same pattern of correlations with I.Q. appears in second grade, with small but significant correlations at midyear giving way to non-significant correlations at the end of first grade. Low correlations could be attributed to the relatively low reliability of the I.Q. measures (the year-to-year correlation between the different tests is .655), the low reliability of teacher's marks, or the relative restriction in range on both variables. Since, however, teachers in the middle-class school are supposed to assign marks with ability partialled out (i.e., by marking each child according to how his performance relates to his own ability), the small or non-significant correlations may also demonstrate that the teachers have successfully implemented the marking policies.

Correlations between I.Q. and marks are much higher in the lower-class school and remain significant throughout the year. For blacks, midyear correlations between I.Q. and marks are considerably higher (by about 0.20) than for whites, but they are not significantly higher. There are no significant correlations between I.Q. and conduct for blacks at either time (+.16 and +.01) but all other I.Q.-Mark correlations are significant. The only statistically significant difference in I.Q. vs. Mark correlations by race occurs for conduct marks at the end of grade one (.01 for blacks vs. .41 for whites).

The differences between sizes of correlations from school to school are striking. The pattern indicates that teachers are effectively implementing the different marking policies in the two schools. In the lower-class school children are supposed to be marked with reference to grade-level performance, and I.Q. is the strongest single predictor of this. Therefore a high correlation between I.Q. and first marks indicates the validity of the teacher's marks given that marking policy.

The effects on children are quite another matter. Average I.Q.'s are not very different in the two schools, but in the middle-class school a child is marked with I.Q. partialled out. As a consequence children can get different rewards in the two places for the same effort. If the middle-class child tries hard, he gets an "A". If the lower-class child tries hard, he still may not get an "A" unless he also happens to be bright. The two policies with respect to marking could lead children to form very different perceptions of efficacy. Another implication of marking policies is that in the lower-class school if the teacher is inept and the child reads below grade-level, the child pays an immediate penalty in terms of

a poor mark. In the middle-class school, an inept teacher may have less damaging effects because the child is not marked so much in terms of his performance, at least in the earliest grades.

Differences in marking practices of first-grade teachers in the middle-class school from one cohort to the next are surprisingly large but nowhere near as large as interschool differences in marking practices. Full assessment of interschool differences, however, must await measurement of intercohort variability in the lower-class school as well.

Children's Expectations

In all three cohorts (two middle-class, one lower-class) correspondence between a child's expectations and his marks on his first report card is at chance levels. Children apparently have no genuine feeling for what marks they will get in any of the three areas.

On the average children's expectations are too high, especially in the lower-class school. A natural question is whether this over-estimation should be taken at face value or whether it comes about for some other reasons, perhaps because children do not understand the task or because they are unable to report accurately, or because they are loath to acknowledge anything less than high expectations to the interviewer.

Two kinds of supplementary evidence bear on the validity of verbally reported expectations especially for middle-class children. First, in one classroom the expectation sampling procedure was carried out on two occasions one week apart. There was a high degree of concordance ($r = .76$) between answers on both occasions. Second, every first-grader in Cohort 2 of the middle-class school and all first-graders in the lower-class school were re-interviewed in depth in June 1973. At that time the interviewer did not ask for expectations at all but rather probed children's understanding of report card by asking questions like the following: What are report cards? What do "Reading" and "Arithmetic" mean? What do the numbers "1", "2", etc. on the report card mean? The purpose was to see whether children's understanding of report cards and of the numerical grading system was clear enough so the interviewer thought they could make a meaningful response in an "expectation interview".*

In the middle-class school report cards are better understood (about 90% of children seemed to have a very good grasp) than in the lower-class school (about 70% of children seemed to have a good grasp). Both groups

---

*At the time of the initial interviews, when children guessed what they would get on their first report card children were asked if they knew meanings of the relevant terms etc. Only after the interviewer was satisfied that the child did understand was the interview continued. Also, of course, the plastic replica of the report card aided in specifying and defining the task. This later probing interview, undertaken by a different interviewer, served as a validity check.

seemed to know what was at issue, however--that report cards were evaluations on how well they were doing in school. It is harder for an interviewer to get lower-class children to verbalize. As already mentioned, reports are issued in kindergarten in the lower-class school so these children have had several more report cards by the end of first grade than middle-class children, although attempts to get lower-class kindergarteners' expectations were not successful. The confusing system of reporting reading progress in first grade in the lower-class school has already been described. This confusion may have interfered with lower-class children's and parents' understanding of the marking system in reading.

In both schools children differentiate between reading and arithmetic in giving their expectations. Expectations for reading exceed those for arithmetic and there is a significant negative correlation between the two areas in both schools. Such differentiation is strong evidence of the validity of children's verbal reports. In both schools the negative correlation between expectations is in sharp contrast to the positive correlation with marks actually awarded in the two areas.

The average level of expectations was higher in the lower-class than in the middle-class school at the middle of first grade in both reading and conduct. Expectations for arithmetic were about the same. There was no difference between expectations of black children and white children in the same (lower-class) school for reading and conduct, but white children were less optimistic about arithmetic than black children.

The significant negative correlation between expectations in reading and in arithmetic in both schools strikes us as the strongest evidence for the validity of children's verbal reports of their expectations. Reliability over time, as in the one-week repeat elicitation, is encouraging but could mean merely consistency on two occasions with no real understanding on the child's part at either time. Significant agreement from midyear to end-of-year can also be taken as an indicant of overtime reliability, but again could represent merely a recall and repetition of what was said earlier rather than a meaningful report.

The average over-optimism in children's expectations in the middle-class school must be interpreted in light of the considerable variability in children's expectations at midyear in first grade. A large percentage of children are over-optimistic but a significant percentage are also over-pessimistic, especially before the first report card. The variability in children's expectations is larger than variability in parents' expectations or in marks awarded by teachers.

The variability of children's initial expectations in the lower-class school is smaller than the variability of both marks and parents' expectations. This relatively small variability in children's expectations results from the extreme skewness (toward the highest mark) that is present in the children's expectations. By the end of first grade more cases of under-optimism begin to appear as expectations decline but expectations still remain, on the average, well above both marks and parents' expectations.

As became clear from further analyses, the level of a child's expectations per se is relatively uninformative in predicting future change. Rather the discrepance between the child's expectations and rated performance is what accounts for change.

The observed lack of racial differences in expectation levels is surprising in some ways but not surprising in others. One might expect small differences between racial groups because the children are still very young. One might also look for small differences, as we did, because this school has been integrated for a number of years and in considerable prior research on other topics in this same school we have found black children and white children to be very similar on other measures (rate of volunteering and the like). When black children are inducted into new or strange situations with white children--as when say children are moved abruptly from a segregated to an integrated setting for experiments--blacks may give evidence of low expectations for themselves (see Cohen and Roper, 1973). These low expectations perhaps reflect the low expectations whites hold for blacks in unfamiliar situations, or the relatively more threatening nature of universities and scientists as perceived by black children. The observations for this research were made in a naturalistic rather than experimental setting, and were rather unobtrusive.

Expectations Over Time

Since expectations generally exceeded marks received at midyear in both schools, one would think that expectations would decrease over the first-grade year. Quite the opposite happens in the middle-class school. Average expectation levels are higher at the end of first grade than at midyear. In the lower-class school expectations generally stayed at the same level, with only a moderate decline for conduct--it still maintained the highest average expectation of the three areas. In the lower-class school there is no difference, furthermore, between blacks and whites in how expectations change over the first-grade year.

Middle-class children get better at anticipating their marks over the first-grade year. In particular they get better at not under-estimating. There is highly significant agreement between children's expectations and the marks they receive in reading at the end of the year (not at midyear). For both arithmetic and conduct agreement approaches significance only at the end of the year. Children seem to learn quickly that teachers give no one the lowest mark and give few children the next-to-lowest, so children whose expectations were low at midyear modify their hopes upward. Children whose expectations were too high at midyear are less apt to modify them downward.

Of those middle-class children whose expectations and marks disagree, the majority have expectations that are too high in reading and conduct, but not in arithmetic. Arithmetic seems to be a topic that generates uncertainty (anxiety?) in first-graders. Teachers do mark harder in it than in other subjects.

Lower-class children do not show significant matching of marks and expectations in any of the three areas at midyear or at the year end. Their high expectations--much higher than middle-class children's--are not modified over the year despite hard marking by teachers--almost 50% get very low marks in reading on the first report card. The most prevalent finding is that expectations remain constant and since most children start with high expectations, they are high also at year's end. Considerable negative feedback seems to be ignored for lower-class children's expectations, if anything, increase rather than decrease. This is perplexing. One explanation might be that the children do not understand what they are doing when they report their expectations to the interviewers. Both the end-of-year depth interviews and the negative association between expectations in reading and arithmetic, however, argue against this simple explanation. Also, results to be discussed in greater detail below show that particular children whose marks increase are recruited from those who did worse than they expected at midyear, and their performance rises to equal expectations. Thus for certain children there is a causal role played by expectations.

The interpretation of the paradox may lie in the impact of day-to-day events rather than the impact of a relatively infrequent event like receipt of a report card. Reading marks were hard to understand (even for the researchers) and reports are given only twice, in November and March. The way reading is taught, though, leads to a high level of individual reinforcement in the classroom administered by both the teacher and the aide. Each spends part of every day in small group sessions with children carrying out exercises related to reading. Children also spend part of every day doing individual written exercises related to reading. In these exercises there is a high rate of positive response by the adults and comments are given in extremely positive terms even if the child's performance has flaws, in order to encourage the child to keep trying. The use of continuous positive feedback, sometimes non-verbal in the form of raisins or M & M candies, is a characteristic of many of the reading programs developed for use with low achievers in recent years. In fact, children who are relatively poor performers may receive more encouragement than better performers, because they get so little self-generated reinforcement from their success at academic tasks that the teacher may need to subsidize them to keep them working. (This possible differential positive reinforcement of poor performers might be invoked as an alternate explanation of why those children whose expectations exceed their marks show a disproportionate likelihood of raising marks but it, unlike an expectation explanation, is unable to explain the other mark phenomena reported on page 5 - 38 and discussed later in this chapter.)

The upshot is that poor-performing children are receiving strong positive feedback face-to-face day after day which contrasts sharply with the negative feedback they get in written form on two occasions during the school year. Persons of any age would be inclined to attend to massive positive feedback especially when the positive feedback is direct, immediate, and unambiguous, and to discount rare pieces of negative feedback. The child's expectations, therefore, may be premised on what he perceives as the everyday behavior of his teacher and little affected by the infrequent and ambiguous written reports. Lower-class children are probably also less able than middle-class children to

appreciate linguistic usages involving mitigating terms (Bernstein, B., 1970) so that the verbal comments teachers make may have a more positive impact than the teacher really intends.

It is hard to imagine how a young child can be brought to learn without considerable encouragement. The middle-class child, however, may be better at encouraging himself when he starts school. For one thing, his lower expectation level will let him experience subjectively-defined success more often. If a child has very high expectations and his performance does not equal them, the contrast leads to negative reinforcement.

It may be dangerous to link heavy encouragement in the classroom with rigid marking practices as is done in the lower-class school. At some point a child must experience severe dissonance when he contrasts his high subjective ideas of his own performance, reinforced by face-to-face interaction with the teacher, with the low marks he gets on report cards. In first grade dissonance may be minimal because marking practices are obscure. In later grades such dissonance is less equivocal--imagine the feelings of the child who is strongly encouraged through the first three years of school when he reviews his report cards and it dawns on him that all his marks have been low or failing.

Marks

In both schools there is significant matching between marks in reading and arithmetic at both midyear and year end. In the middle-class school conduct marks show significant matching with both reading and arithmetic marks at all times during the first grade but the percentage of cases matching in these areas declines from midyear to year end. The lower-class school generally shows no relationship between conduct marks and the academic areas, an exception being arithmetic at year end where a low level of significance is attained.

This general pattern is not too surprising. Marks in the academic areas might be expected to be related whether marking was done in terms of grade norms (as in the lower-class school) or in terms of "effort" (as in the middle-class school). Marking in terms of grade norms should produce conduct marks that are reasonably independent of academic marks--and this happens in the lower-class school. Marking the academic areas in terms of effort might be expected to initially produce marks that show a relation to conduct marks as the various behavioral manifestations of effort available to teachers for marking purposes seem closely related to classroom behavior or children's general attempts to behave. The observed relationship between arithmetic and conduct marks is of a halo effect in middle-class teachers' marking, for there is much more correspondence between arithmetic and reading than would be expected with I.Q. effectively partialled out. It also seems reasonable that as the year progresses teachers should show increased ability to separate children's academic efforts from their conduct efforts.

The variable correlations between I.Q. and marks from the middle of the year to the end of the year and between schools could also directly reflect the different marking policies in force at the two schools. In the middle-class school marks are supposed to be awarded in terms of the child's living up to his ability level whatever that level is. Pragmatically a measure of teachers' effective implementation of this policy are the weak correlations of I.Q. with midyear marks and the lack of such correlations at year's end in both first and second grades at the middle-class school. Likewise the significant correlations between I.Q. and marks in the lower-class school can be seen as reflecting teachers' success at rating children in terms of grade-level performance. Also, of course, the relatively greater variability of both I.Q. scores and marks in the lower-class school favors the appearance of higher correlations between I.Q. and marks there in contrast to the middle-class school.

The data shed light on marking practices per se. It is, first of all, surprising how much difference there is in marking distributions from one year to the next in the middle-class school. It is also surprising that so many very poor marks (up to 50%) are given in the lower-class school on the first report card. The child who is receiving his or her first marks processes this feedback in ways that have effects on performance as the data show for middle-class children, and in ways that seem to have no effect for lower-class children. This lack of effect could imply an "insulation effect"--when feedback is negative the child insulates himself against it. Such insulation may have advantages for preserving the self-image but it may be deleterious in terms of monitoring and thereby improving performance. The implications of the two different marking systems found in the two schools--one in relation to ability and the other in relation to grade norms--may be profound in terms of long-range performance. Our data show lower-class children's expectations were maintained at very high levels to the end of grade one. It will be interesting to trace the course of these expectations in subsequent years as is planned for future work on this same research project.

A number of comments about marking practices are in order. First, teachers are asked to evaluate the student's performance in 27 areas in the middle-class school and in 22 areas in the lower-class school, surely a difficult task with 25 or 30 individual children to be rated. As far as first-graders are concerned, our research suggests that most have a fair understanding of what is being rated in reading, arithmetic and conduct, but children do not understand other areas, such as spelling. Teachers themselves, when queried as to marking practices, show that they operate using different criteria and also that they themselves do not have very clear cut standards for what a topic like "Language" consists of.

Teachers treat the sexes much the same at the start of school but there are trends for girls to get better marks in reading and conduct by the end of first grade.

It is impossible to tell what the overall effect of the lower-class school's marking practices will be, but as already suggested, uniformly poor marks may lead the child to disregard marks or to revise his expectations drastically (the latter has not happened by the end of first grade).

In the middle-class school marks are so high that there is little room for improvement, especially after final marks are given in grade one. Since it has been shown that improvement in marks is possible if expectations exceed them, by creating a ceiling in marks teachers could be creating a ceiling for achievement.

For middle-class children, who rather uniformly get high marks at first, there may be long-run problems, first in the later grades of elementary school when child-to-child comparisons begin and then later in junior high and high school when competition increases. Having expectations that are very high can lead to improved performance early but it could lead to adjustment problems later when the students finds himself up against stiffer competition.

Discrepances Between Marks and Expectations

The main purpose of this research is to uncover causal factors: do children's (or parents') expectations causally impact on performance, and if so, how? Looking at average expectation levels one gets little feeling for how causal effects may occur. The following sections address the issue of causal sequences and examine in more detail what happens to children sequentially.

One might think that children with high marks are more astute and that they therefore are better at forecasting their marks. However when expectation-mark agreement is stratified by mark received at midyear, middle-class children of all performance levels seem to formulate their expectations in the same way. Children receiving A's and those receiving D's show approximately the same distribution of expectations (as do those in between).

Another reasonable hypothesis might be that children with low expectations, more than children with high expectations, would manifest a high amount of agreement between forecasts and performance because a poor perfromance is under the child's control. But middle-class children who forecast low marks, and they are relatively few in number, get about the same percentage of low marks as the class as a whole so this line of reasoning also does not hold.

For lower-class children the data do not permit a meaningful analysis of expectation-mark agreement at first report card in terms of mark or expectation level because, as remarked in Chapter 6, the combination of exceedingly high child expectations with relatively low mark assignments severely restricts possible patterns of outcomes. Also with data for only a single cohort the case base is small.

By the end of first grade in reading there is increased expectation-mark agreement for middle-class children (increasing from about 55% to 80%). Change in marks and expectations seems related <u>not</u> to simple performance level <u>or</u> to simple expectation level, as already stated, but rather to the difference between performance and expectations. There are too few

data for the lower-class school to allow definitive analyses, but in both arithmetic and conduct, those children whose marks do improve are disproportionately recruited from among those who earlier showed disagreement between mark and expectations. There is a tendency for both marks and expectations to change in a way to reduce expectation-mark discrepance.

Feedback effects will now be discussed for the middle-class school only. If a child did worse than he expected at midyear, his expectations at year's end tend to remain the same and his mark is brought into line. If he did better than he expected, he is very likely to modify his expectations upward and keep the mark the same. There is a consistent observation as well of the following types of cases: a child's mark remaining the same if his mark equalled what he expected, and a child's mark going down if he did better than he expected.

The majority of children whether their midyear mark exceeded, equalled or was below their expectation can expect the same mark at year's end that they received at midyear, because there is great consistency in marks awarded. There is, nevertheless, a consistent (and significant for reading and conduct) observation of cases where marks move up, down, or remain stationary to agree with expectations. The data suggest that students' expectations causally affect the marks they receive even if the degree of association is rather low.

Finding a causal role for expectations is one of the major achievements of this research. Only with data aggregated across two (middle-class) cohorts to supply an adequate number of cases for analysis could this effect emerge.

The data show that a child's cognitive state, his estimate of his own performance, is instrumental in changing that level of performance. Unfortunately the data in the lower-class school are not yet extensive enough to allow an investigation of the same sort, but what data are there, as already stated, point to the same finding.

The other side of the coin must not be lost sight of, however. Of all the children followed, the most likely outcome is for both expectations and marks to remain the same. In the middle-class school where almost all children are awarded A's and B's and where expectation levels are high but not inordinately high considering the high average performance level, the stability of expectations leads to high expectation but not to an unrealistic cognitive state. In the lower-class school, where expectations also persist at high levels, if future analyses bear out present tendencies, several outcomes are possible. First the child may insulate himself from negative feedback--he may block out awareness of, or dissonance over, low marks--and so be hindered in monitoring his performance. Second, if there are downward shifts in expectations from time to time as the child becomes aware of low-level performance, serious limits could be imposed thereafter on aspirations.

We have shown so far that young children's expectations are high, resilient, and responsive mainly to positive feedback. More data are needed to see whether, if expectations are lowered, they are also resilient to change thereafter.

Parents' Expectations

Parents and children do not agree on expectations in either school. Parents' expectations in the middle-class school show highly significant agreement with teachers' marks in reading at midyear, and with marks in arithmetic and conduct both at midyear and end of year. Parents' and teachers' marginals tend to correspond in the middle-class school. Parents tend to "play it safe", however, and most forecast a "B" in all areas.

Parents do seem able to identify those children who will perform poorly or very well. One might think that a superior performer would be more easily identified, and this is true even at the first report card. There are significant trends toward increased agreement between parents forecasts and year-end marks in arithmetic and conduct but not in reading. The pattern of this increased agreement is consistent with a causal hypothesis viewing parents' expectations as the independent variable. Parents from one year to the next in the middle-class school show a higher degree of consistency in their expectations than that shown by either teachers marks or children's expectations.

When middle-class parents' expectations are not correct, they are in the direction of under-estimating in reading and arithmetic. There is a high correlation between parents' expectations in reading and arithmetic but no relation between either of these and conduct.

Middle-class parents here demonstrate a compartmentalization of expectations for their children's school performance that Melvin Kohn (1969) would predict. They do not see classroom behavior--ability to sit still, to be docile, to be deferent--as related to intellectual performance.

In contrast to trends noted for middle-class parents, the parents of lower-class children do not have expectations that are at all veridical with first marks in reading. They do match marks in reading at the end of first grade, however. In the lower-class school the preponderance of low marks in reading at midyear is not anticipated by parents. Lower-class parents also over-estimate children's performance in arithmetic at midyear and agreement with marks in arithmetic is not significant at year's end.

Lower-class parents expect higher marks than their children receive. Interestingly, on the average, their expectations are a little lower than middle-class parents' in reading and arithmetic but are about equal in conduct. There is a major difference between marking practices in the lower-class and middle-class schools as already noted. The actual distributions used by teachers in the middle-class school agree well with what both sets of parents anticipate. The low marks awarded in the lower-class school lead to parents there over-estimating their child's performance, whereas in the middle-class school there is under-estimation, if anything.

Parents' involvement with school, or the attention parents give to school matters, might be indexed by the levels of response noted in this research from the two sets of parents. In the middle-class school we were able to attain a response rate of 86% using a mail questionnaire when we

sought parents' expectations. For Cohort 1 in first grade we were able to see 92% of the parents when they visited school during American Education week. Using a trained interviewer to contact lower-class parents, preceded by a flyer carried home by the child and persisting through three call-backs, we were able to secure responses from 74% of lower-class parents. When we attempted to interview lower-class parents by seeing those who came to school in American Education week, our response rate was less than 10%.

One set of parents thus seemed in close touch with school, the other set not. In the middle-class school the PTA actually aided the researchers in contacting parents. The PTA in the lower-class school is inactive. Communication fluency may be what causes middle-class parents to have more realistic expectations for a child's performance. By communicating with other parents and with teachers, middle-class parents may get enough information to be accurate.

Sex Differences

Reading marks in both schools do not show a significant association with sex at midyear in either school or at year end in the lower-class school. By year's end, however, more boys in the middle-class school get the poorer marks. At no time is there any significant association between marks in arithmetic and sex in either school.

In conduct there are large sex differences at midyear and at year's end favoring girls in the lower-class school. In the middle-class school there is a strong trend for the lowest conduct marks to go to boys. There is a slight tendency for middle-class parents to expect poor conduct from boys as compared to girls, and surprisingly, lower-class parents expect the same performance in conduct for boys and girls. The expectations of middle- and lower-class parents for boys and girls are at odds with what might have been predicted.

Racial Differences

Results in terms of racial breakdowns are not extensive because they are based on only one cohort. Some preliminary and tentative findings with respect to race are nevertheless worth summarizing.

In first marks, there is no association between race and marks in reading, although both in arithmetic and conduct white children's marks tend to be higer (significantly so for conduct). The differences are very small however by practical criteria, less than one-quarter of a unit in both areas.

At the end of first grade there is movement toward whites getting higher marks in reading (2.43 vs. 2.69) but the improvement in marks of both

groups over the year is much more striking than the difference between them at year's end. Whites' average improvement is 0.71 grade points and blacks' average improvement is 0.47 grade points. In arithmetic and conduct the difference by race is about the same as at midyear and only slightly favors whites.

Children of both races expect the same marks. Parents of both races have similar expectations in reading and arithmetic but white parents expect better conduct marks (1.61 vs. 2.08). Both sets of parents have expectations which are not consistent with the marks their children receive. There is some limited indication that white children are more cognizant of earlier marks in shaping their year-end expectations than black children.

Two teachers are white and two black and each class has approximately 40% white and 60% black children. As far as these data shed light on teacher differences, there do not appear to be any differences associated with race for teachers.

Implications and Future Research

More work is needed both to extend the time span over which particular children are followed and also to increase the number of first-graders for study. Many more analyses, and more searching analyses in terms of causal hypotheses, were possible for middle-class children because two cohorts could be aggregated. An even larger sample is needed for the lower-class school because there one wishes to carry out racial breakdowns as well. Important measures still missing are standardized test scores for reading and arithmetic. (These tests are not given in first grade.) They represent a criterion of performance badly needed as a dependent variable in our analysis. Teachers' marks are compounded of expectations and judgment related to I.Q. and other factors, making such marks of great interest from the standpoint of Pygmalion effects or of interpersonal influences, but they are not very good barometers of performance. In fact, major future interest will attach to the relations between expectations, teachers' marks and standardized achievement test scores.

To date some significant findings are available, however. Children at the start of school have a great deal of optimism about their ability to perform. Their expectations for the most part are high, particularly for reading. This suggests that the "protective circle of the family", whether middle-class or lower-class, nurtures children who are self-confident and optimistic when they start school. The expectations of parents further supports this claim, for both middle-class and lower-class parents are on the whole optimistic about their first grader's likely success in school, even though lower-class parents are not quite as optimistic as middle-class parents. What is somewhat perplexing is how lower-class parents can express such confidence if they have older children whose performance has been poor, as must often be the case.

Parents' expectations are remarkably constant in the middle-class school and show significant agreement with children's marks in all three areas. Lower-class parents' expectations agree with marks in only one instance, at the end of first grade in reading, after children's reading marks have improved noticeably. Data are limited but changes in marks in the middle-class school seem more closely attuned to discrepances between parents' expectation and marks than discrepances between children's expectations and marks. (The possibility that joint discrepances of the same sort by both parents and children are occurring has yet to be studied.) Add to this the fact just noted that lower-class parents' expectations match marks at the end of first grade in reading but not at the middle of the year, and the inference is that parents' are influential in cases of children whose marks move up.

There is, however, a question clearly raised at this time which can be answered only with future data. It is that movement in marks is associated with a prior discrepance between expectations and marks such that differences between the two are abolished--they move toward consistency. If the middle-class child did better than expected, his mark tends to drop. If he did worse than expected his mark tends to rise, but even if his mark does not rise his expectations are more likely to remain the same than drop.

In the lower-class school no one did better than he expected. Marks are much lower and there is a significant tendency for children whose marks in arithmetic and conduct improve over the year to be disproportionately recruited from those whose expectations exceeded their first semester mark. This statistical significance, however, is linked to an effect that is rather small in absolute size. (In reading a similar trend is not significant but is worth attention because it is consistent with the trends in arithmetic and conduct.)

Improvement in marks in both schools, then, is related to earlier mark expectation discrepances but improvement in marks occurs for only a minority of children in either school. Expectations move more easily than marks and also in a way to reduce expectation-mark discrepance, but are characterized by a bouyancy effect. They tend to rise more easily than they tend to fall. Lower-class children's expectations do not fall, and their expectations are very high to start with. A disequilibrium between marks and expectations is noticeable at midyear first grade and persists over the year because marking is "hard" and expectations are extremely high. What happens as time goes on? It would be difficult for marks to keep improving when children are marked in terms of grade-level norms and yet so far there has not been a decline in expectations. Do expectations drop later? One would think they would have started to drop already but they have not.

In an earlier review of reading models and language socialization one of the authors (Entwisle, 1971) noted "evidence is accumulating that socioeconomic status is a more crucial influence on reading performance than I.Q." The present report looks not at linguistic factors or cognitive habits, as the earlier paper did, but looks at social factors or organizational (school) variables. The same conclusion can be stated here, however, with even more conviction. The average difference in I.Q. is

not large between schools or between races but reading and arithmetic performance by first graders differs by race and school, and all prior experience suggests that there will be significant differences on standardized tests in reading and arithmetic between schools when these children are tested in second grade. Informal classroom observation reveals very large differences in reading proficiency at the end of first grade.

We intend to investigate the present data set much more thoroughly to gain insight into what may be responsible for these differences. Parents of lower-class children may be setting themselves up to administer negative reinforcement--by having higher expectations than their children can deliver they may be boxing themselves into a corner of not being able to be pleased with the child's performance. By marking on an absolute scale the teachers in the lower-class school may also be removing their ability to provide positive reinforcement. If a child is trying hard and sees no obvious result, i.e., no high mark, his effort must slack off.

We suspect that forces are already at work during first grade in addition or prior to any reinforcement effects. One of these is the effective time the child is coupled into the learning system. We have yet to analyze absence data which may be used as an indicator of "effective time". We have noted informally, however, a much higher absentee rate in the lower-class school. We also have noted much less of children's and teachers' time spent in cognitive activities because children are late or not prepared for school (unsuitably dressed and the like). To this one can add, although we have as yet no measure of it, the likelihood that the lower-class children are coming to school having had a poor breakfast or no breakfast at all. We intend to investigate all these possibilities as the research continues.

BIBLIOGRAPHY

Adelman, H.S. Reinforcing effects of adult non-reaction on expectancy of under achieving boys. Child Development, 1969, 40, 111-122.

Alexander, Karl, & Eckland, Bruce K. Effects of education on the social mobility of high school sophomores fifteen years later (1955-1970). Final Report to National Institute of Education, Project No. 10202, 1973.

Bernstein, B. A sociolinguistic approach to socialization: with some reference to educability. In F. Williams (Ed.), Language and Poverty. Chicago: Markham Publishing Company, 1970. Pp. 25-61.

Bloom, B.H. Stability and Change in Human Characteristics. New York: Wiley, 1964.

Bronfenbrenner, U. An emergency theoretical perspective for research in human development. Mimeo. Cornell University, Spring, 1973.

Brookover, W.B., Erickson, E.L., & Joiner, L.M. Self-concept of ability and school achievement, III. U.S. Office of Education, Cooperative Research Project No. 2831, East Lansing: Office of Research and Publications, Michigan State University, 1967.

Brookover, W.B., LePere, J., Hamachek, D., Thomas, S., & Erickson, E. Self-concept of ability and school achievement, II. U.S. Office of Education Cooperative Research Project No. 1636, East Lansing: Office of Research and Publications, Michigan State University, 1965.

Brookover, W.B., Patterson, A., & Shailer, J. Self-concept of ability and school achievement. U.S. Office of Education, Cooperative Research Project No. 845. East Lansing: Office of Research and Publications, Michigan State University, 1962.

Brookover, W.B., Thomas, S., Patterson, A. Self-concept of ability and school achievement. Sociology of Education, 1964, 37, 271-78.

Brophy, J.E. & Good, T.L. Teachers communication of differential expectations for children's classroom performance. The Journal of Educational Psychology, 1970, 61, 367-374.

Cohen, D.K., Pettigrew, T.F., & Riley, R.T. Race and outcomes of schooling. In F. Mosteller and D.P. Moynihan (Eds.), On Equality of Educational Opportunity. New York: Random House, 1972. Pp. 343-370.

Cohen, E.G. Interracial interaction disability. Human Relations, 1972, 25, 9-24.

Cohen, E.G., & Roper, S. Modification of interracial interaction disability. American Sociological Review, 1973, 37, 643-647.

Coleman, J.S., et al. Equality of Educational Opportunity. Washington, D.C.: U.S. Government Printing Office, 1966.

Dickstein, E. The development of self esteem: theory and measurement. Unpublished doctoral dissertation. The Johns Hopkins University, 1972.

Entwisle, D.R. Developmental sociolinguistics: inner-city children. The American Journal of Sociology, 1968, 74, 37-49.

Entwisle, D.R., Grafstein, D., Kervin, J., & Rivkin, M. Giant steps: a game to enhance semantic development of verbs. Report No. 81, Center for Study of Social Organization of Schools, Baltimore, Md. The Johns Hopkins University. 1970.

Entwisle, D.R., & Webster, M. Raising children's performance expectations, Social Science Research, 1972, 1, 147-58.

Entwisle, D.R., & Webster, M. Research notes: Status factors in expectation raising. Sociology of Education, 1973, 46, 115-126.

Entwisle, D.R., & Webster, M. Raising children's expectations for their own performance: A classroom application. Chapter 7 in J. Berger, T. Conner, & M.H. Fisek (Eds.), Expectation States Theory: A Theoretical Research Program. Cambridge: Winthrop Publishers, 1974a.

Entwisle, D.R., & Webster, M. Expectations in mixed racial groups, Sociology of Education, 1974b, in press.

Entwisle, D.R., & Webster, M. Raising expectations indirectly. Unpublished manuscript, The Johns Hopkins University, 1974c.

Finn, J.D. Expectations and the educational environment. Review of Educational Research, 1972, 42, 387-409.

Gibby, R.G., & Gibby, R.G. Jr. The effect of stress resulting from academic failure. Journal of Clinical Psychology, 23, 35-37, 1967.

Hamburg, D.A. Emotions in the Perspective of Human Evolution. In Expression Of The Emotions of Man, Peter H. Knapp (Ed.). New York: International Universities Press, 1963.

Husen, T. Talent Opportunity and Career. Stockholm: Almquist & Wilsell, 1969.

Kagan, J., & Moss, H.A. Birth to Maturity. New York: Wiley, 1962.

Katz, I. Experimental studies in Negro-White relationships. In L.Berkowitz (Ed.), Advances in Experimental Social Psychology, Volume 5, New York: Academic Press, 1970. Pp. 71-118.

Katz, I., & Benjamin, L. Effects of white authoritarianism in biracial work groups. Journal of Abnormal and Social Psychology, 1960, 61, 448-556.

Katz, I., Goldston, J., & Benjamin, L. Behavior and productivity in biracial work groups. Human Relations, 1958, 11, 123-41.

Kerensky, V.M. Reported self-concept in relation to academic achievement in an inner-city setting. Dissertation Abstracts, Wayne State University, 1966.

Kohn, M. Class and Conformity. Homewood, Illinois: The Dorsey Press, 1969.

Kraus, P.E. Yesterday's Children. New York: Wiley, 1973.

Lambert, N. Paired associate learning, social status, and tests of logical concrete behavior as univariate and multivariate predictors of first-grade reading achievement. American Educational Research Journal, 1970, 7, 511-528.

Lamy, M.W. Relation of self-perceptions of early primary children to achievement in reading. In J.J. Gordon (Ed.), Human Development Readings in Research, Chicago: Scott, Foresman, 1965. P. 251.

Lesser, G.S. Learning, teaching, and television production for children: the experience of Sesame Street. Harvard Educational Review, 1972, 42, 232-272.

Morse, R. Self-concept of ability and school achievement: a comparative study of Negro and Caucasian students. In W.B. Brookover, E.L. Erickson, & L.M. Joiner, Self-Concept of Ability and School Achievement III, Final Report, Project No. 2831, Michigan State University, February, 1967.

Oden, M.H. The fulfillment of promise: 40-year follow-up of the Terman gifted group. Genetic Psychology Monograph, 1968, 77, 3-93.

Palardy, J.M. What teachers believe--what children achieve. Elementary School Journal, 1969, 69, 370-374.

Porter, J. Black Child, White Child, Cambridge, Mass.: Harvard University Press, 1971.

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

Rosenberg, M., & Simmons, R.A. Black and White Self-Esteem. Washington, D.C.: American Sociological Association Rose Monograph Series, 1971.

Rosenthal, R., & Rubin, D. Pygmalion reaffirmed. Mimeo, Department of Social Relations, Harvard University, July 1971, Prepublication version of an invited contribution to the volume Pygmalian Reconsidered, National Society for the Study of Education.

Seaver, W.B., Effects of naturally induced teacher expectancies. Journal of Personality and Social Psychology, 1973, 28, 333-342.

Shaw, M.C., Edson, K., & Bell, H. The Self-concept of bright underachieving high school students as revealed by an objective check list. Personnel and Guidance Journal, 1960, 39, 193-96.

Snedecor, G., & Cochran, W.A. Statistical Methods. (Sixth ed.). Ames, Iowa: Iowa State University Press, 1967.

Wattenberg, W.W., & Clifford, C. Relation of self-concepts to beginning achievement in reading. Child Development, 1964, 35, 461-67.

Webster, M., & Entwisle, D.R. Expectation effects on performance evaluations. Unpublished manuscript in preparation, The Johns Hopkins University, 1974.

Webster, M., & Sobieszek, B. Sources of evaluations and expectation states. Chapter 5 in J. Berger, T. Conner, & M.H. Fisek (Eds.), Expectation States Theory: A Theoretical Research Program. Cambridge: Winthrop Publishers, 1974a.

Webster, M., & Sobieszek, B. Sources of Self-Evaluation: A Formal Theory of Significant Others and Social Influence. New York: John Wiley, 1974b.

Williams, F. Psychological correlates of speech characteristics: on sounding disadvantaged. Journal of Speech and Hearing Research, 1970, 13, 472-488.