Doctoral Dissertations 1896 - February 2014

1-1-1991

Perceptions of selected upper elementary public school teachers toward grouping and evaluation of pupil learning progress in reading.
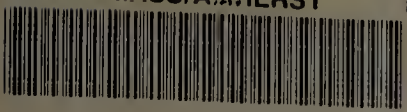
Kriner Cash
University of Massachusetts Amherst

Recommended Citation

Cash, Kriner, "Perceptions of selected upper elementary public school teachers toward grouping and evaluation of pupil learning progress in reading." (1991). *Doctoral Dissertations 1896 - February 2014*. 4734.
https://scholarworks.umass.edu/dissertations_1/4734

PERCEPTIONS OF SELECTED UPPER ELEMENTARY PUBLIC

SCHOOL TEACHERS TOWARD GROUPING AND EVALUATION

OF PUPIL LEARNING PROGRESS IN READING

A Dissertation Presented

by

KRINER CASH

Submitted to the Graduate School of the

University of Massachusetts in partial fulfillment

of the requirements for the degree of

DOCTOR OF EDUCATION

May 1991

School of Education

PERCEPTIONS OF SELECTED UPPER ELEMENTARY PUBLIC SCHOOL TEACHERS TOWARD GROUPING AND EVALUATION OF PUPIL LEARNING PROGRESS IN READING

A Dissertation Presented

by

KRINER CASH

Approved as to style and content by:

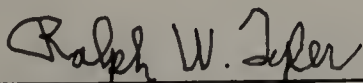_Robert L. Sinclair_
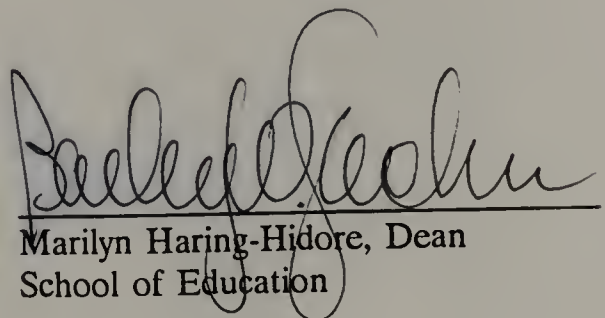Robert L. Sinclair, Chairperson

David M. Bloome
David M. Bloome, Member

E. Ann Sheridan
E. Ann Sheridan, Member

Ralph W. Tyler
Ralph W. Tyler, Member

Marilyn Haring-Hidore
Marilyn Haring-Hidore, Dean
School of Education

ACKNOWLEDGEMENTS

Robert Sinclair has cultivated for his students the kind of learning environment that is intensely stimulating, accepting, and appreciative of many different individuals and cultures. It is a place where people care about each other and support each other. Bob is demanding yet patient with each of us. He has provided time and opportunities for us to learn. The ability to develop and nurture many differing personalities for effective leadership in education is a rare skill. Bob Sinclair is a master mentor.

I appreciate and gratefully acknowledge the contributions of Dr. Ann Sheridan and Dr. David Bloome to the improvement of this study.

I am grateful to Ralph Tyler for sharing his vast wisdom and professional insights with me in different places and settings.

I acknowledge the importance in helping me complete this work that the love and support of members of my extended family and friends have had. You are there for me - across the miles and through the years.

And three special dedications:

To my parents Gene and Donna Cash. You laid a rich foundation and throughout your lives gave me all the love and encouragement that I needed, whenever I needed it.

To my three sons, Kofi, Asil, and Jade. You have given me unconditional love and reinforced for me the meaning of persistence and perspective in my work. A significant part of this accomplishment is for you.

Finally, but not last, this dissertation is dedicated to my wife, Lisa. I have learned and benefitted from your approach to people and life. We are one in Love and Spirit.

ABSTRACT

PERCEPTIONS OF SELECTED UPPER ELEMENTARY PUBLIC

SCHOOL TEACHERS TOWARD GROUPING AND EVALUATION

OF PUPIL LEARNING PROGRESS IN READING

MAY, 1991

KRINER CASH, B.A., PRINCETON UNIVERSITY

M.A., STANFORD UNIVERSITY

Ed.D., UNIVERSITY OF MASSACHUSETTS

Directed by: Professor Robert L. Sinclair

This descriptive study examined the perceptions of teachers toward grouping and evaluation of pupil learning progress in reading. Instructional grouping decisions by teachers were viewed as part of the evaluation function of teaching. Teacher practices in grouping and evaluation were assumed to represent significant classroom level indicators of equity and excellence in the education provided for children. A major priority for the inquiry was to discover whether the evaluation information collected about students lead teachers to change the practice of ability grouping.

Three major research questions guided the study:

1.   What criteria do teachers report they use to group pupils for instruction in reading?

2.   What are teachers' perceptions of the similarities and differences in ways that they evaluate pupil learning progress across instructional groups for reading?

3.   What are teachers' perceptions of ways they use data from their evaluation of pupil learning progress in reading?

The study adapts assumptions from interpretive and critical theoretical perspectives complemented by a qualitative research design to describe the perceptions of teachers for grouping and evaluating students in reading. Data are drawn from interviews with 23 experienced teachers representing grades 5 and 6 in ten schools in western Massachusetts.

Findings indicate that teachers tend to rely on previous teacher recommendations and the results of their own informal assessments as criteria for grouping students by ability within the classroom. Teachers assign students to within-class ability groups during the first few weeks of school. Students tend to remain in the reading groups to which they were initially assigned. Teachers' evaluation methods and performance criteria vary depending on whether they are in a basal or non-basal reading curriculum. Teachers reported that they expect students placed in higher ability groups to proceed at a faster pace through the reading curriculum and to produce written work that was more detailed and of higher quality than students placed in lower ability groups. Teachers report numerous uses of data from pupil evaluation. The primary uses are for communicating with parents and students and for improving instruction. The study concludes that despite the diverse ways that teachers evaluate student learning progress in Reading, little student mobility occurs across the instructional groups, nor does the evaluation data collected by teachers suggest to them a need to change the practice of ability grouping.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

xi

CHAPTER I

OVERVIEW OF THE STUDY

The purpose of this chapter is to provide an overview of the study. The research problem, purpose, key terms, significance, and delimitations of the study are discussed in turn.

Statement Of The Problem

In at least every decade since 1890, significant debate has been aroused by those concerned with the quality of American public education. A particular perspective of the national problem followed by proposals for reform most often characterize the substance of these debates about education. Consideration of school reforms over the last 100 years suggests that reform movements tend to move in cycles, shifting in emphases between equality and excellence (e.g., Tyack, 1974; Kirst, 1984; Tyler, 1987). For example, the vigorous "educational excellence" rhetoric rejoined in the early 1980s is now being translated into policy at state and institutional decision-making levels. The emphasis this time has focused on quality while issues of equity have been virtually ignored. As practitioners scurry to interpret and adopt recommended changes, it is crucial that educators keep close vigilance over new regulations with a view toward insuring a proper balance in achieving increased standards of excellence and equality of educational opportunity for all children of all families. There is no sensible reason to continually fluctuate from one ideal to the other when both can be profitably blended to guide our course for even better education.

The twin principles of equality and excellence are complex ones. There are many significant factors that affect the ability of an enterprise as large as American

public education to deliver on these two cherished democratic principles at various levels of curriculum decision-making.[1]

For example, at the societal level, conflicting economic values, conflicting social philosophies, and conflicting political interests are continually vying for attention to determine the purpose and substance of education for our children (Kirst, 1984). At the classroom level, the teacher's academic preparation, professional knowledge and experience, beliefs, values, prejudices, and personal expectations for the profession (e.g. Connelly and Elbaz, 1980; Barr and Dreeban, 1983) are some of the variables which, when interacting together, have a tremendous influence over how excellence and equity will be defined for the learner in school.

It follows, then, that teachers can be viewed as one significant force in implementing reform at the school or institutional level (Tyler, 1987). In past cycles of reform, however, teachers' agenda were often excluded or placed backstage. This was an unwise policy since school reform movements essentially boil down to getting teachers and students to become more motivated about doing well in school. Responsibility for educational improvement ultimately resides with teachers. In fact, the amazing stability of certain teaching behaviors decade after decade despite mighty efforts to alter these behaviors (Cuban, 1984) would suggest that for reform movements to be successful they must make sense to teachers. The rhetoric of the present educational reform effort places teachers appropriately at center stage, (read "teacher empowerment," "shared decision-making," etc.). It is unclear, however, how far this language has translated into meaningful practice at school and district levels.

In short, while rhetoric about the meaning of excellence and equity has flared heatedly at the level of national debate among idealogues and scholars, the meaning

perspectives of teachers have remained peripheral to these debates and are glaring in their omission from conventional research studies on teaching upon which most of the prescriptive claims of the 1983 to present reform effort are based. Hence, if the central issue for educational research and reform centers on educational improvement, then the meanings that teachers attach to significant improvement in teaching will have to receive far more deliberate attention. This study attempts to continue addressing this need by bringing to the fore of the inquiry the meaning that selected issues in curriculum decision-making that have implications for equity and excellence have from the viewpoint of the teacher-participant. Specifically, the present study proceeds at the classroom level of curriculum decision-making by investigating the perceptions of teachers toward two dimensions of learner access and success - 1) criteria for grouping students for instruction; and, 2) the evaluation of pupil learning in instructional groups for reading.

Reading is used in the present study as a means for examining the issue of school equity because it is one subject traditionally given primacy in the elementary curriculum. Reading is also notorious as one subject in which students are consistently grouped according to ability (e.g., Borg, 1965; Allington, 1983; Hiebert, 1983).

The primary objective of this study was to examine and interpret how selected upper elementary public school teachers say they evaluate the learning progress that students are making in instructional groups created to teach Reading. A major priority for the study was to discover whether or not the condition of grouping students by ability prevails in the environments for learning that teachers construct for and with students. In light of our expressed concerns for equity, this questionable educational practice must continue to be examined carefully by teachers employing its use and by scholars who desire to replace ability grouping with more compelling instructional

3

practices. Through this investigation, it was interesting to find, however, that teachers do not call into question the practice of grouping students by ability, even when many of their pupils were not performing well in reading and writing.

For classroom teachers utilizing the structure of ability groups for Reading instruction, there seem to be at least three thorny pupil evaluation problems. First, classroom teachers employing ability groups must use reasonably valid means to assign students to these groups. That is, they must properly diagnose each student's reading ability before instruction begins so that an initial determination can be made about which students will go into which groups within the classroom. Second, teachers must use appropriate means to evaluate how well students are advancing their reading skills and forming constructive attitudes toward reading. Third, teachers must decide whether and how to use information yielded from their evaluations to alter instruction for students in the existing ability groups.

The tasks associated with these student evaluation problems can be incredibly complex and demanding for the classroom teacher. Some research studies on classroom assessment have found that as much as 40% of a teacher's professional time may be spent directly involved in assessment-related activities (Stiggins, 1988):[2]

> ...This includes time spent designing, developing, selecting, administering, scoring, recording, reporting, evaluating, and revising such items as daily assignments, tests, quizzes, observations and judgments about student performance, and oral question-and-answer sessions. (p. 364)

Once again, pupil evaluation tasks may be amplified in Reading, since in this subject teachers are traditionally managing three or more groups simultaneously and are often evaluating and using data for making instructional decisions.

Although a great deal of time and energy may be spent in teacher-directed assessments of how well students are performing in the classroom, it is not clear how prepared teachers are to handle the evaluation aspects of teaching nor how they use the information they obtain from student evaluation. Only modest evidence is available that reveals how teachers actually evaluate student learning and make judgments about their progress (e.g. Stiggins et al., 1986). Thus, in addition to teacher responses to the major research questions that guide the study, this investigation also provides insight into some of the emergent concerns, anxieties, and professional development needs of teachers around the issues of grouping and evaluating student learning.[3]

In sum, if schools are to create equal opportunities for students to learn, as education in a democratic society presupposes, then teacher perceptions of their practices in a prevalent condition of ability grouping can be viewed as symbolic of the modern school's application of the tenet of equality. And, if teachers have the responsibility to help all students learn at high levels of accomplishment, then teacher perceptions of the methods they employ to determine the progress of students in learning to read while in groups arranged by ability can be viewed as symbolic of the modern school's application of the tenet of excellence. More studies that move away from traditional research perspectives and adopt interpretive and critical science perspectives are needed so that we can better understand the points of view of teachers toward these essential tenets. Ultimately, the goal for this researcher and others would be to free teachers and students from any and all conditions that repress or hinder learning.

Purpose Of The Study

The purpose of the study is to describe and analyze the perceptions of selected fifth and sixth grade elementary public school teachers toward grouping and evaluation of students for reading instruction. The criteria teachers use to group students for reading instruction, the means teachers use to evaluate reading progress for individuals in the various groups, and the decisions teachers make for using the information they obtain from their pupil evaluations are seen as important interrelated dimensions of this purpose. Three major research questions, analogous to the three parts of the purpose described above, defined the parameters of the study:

Research Question 1: What Criteria do Teachers Report that they Use to Group Pupils for Instruction in Reading?

Research Question 2: What are Teachers' Perceptions of the Similarities and Differences in Ways that they Evaluate Pupil Learning Progress across Instructional Groups for Reading?

Research Question 3: What are Teachers' Perceptions of Ways they Use Data from their Evaluation of Pupil Learning Progress in Reading?

The investigator used qualitative interview methods to uncover what teachers perceived they do in practice when they evaluate pupil learning in Reading. The present study contributed to important descriptive work being done on academic grouping as it relates to the larger issue of equity and excellence in teaching (see, for example, Trimble, 1988; Harrison, 1989).[4] To enhance the potential generalizability within the body of data of the study, schools selected for the research included those which reflected diversity in teacher and student populations and were demographically different in size, age, and geographic location.

6

This section defines two complex terms explicitly and implicitly referred to in the present study - evaluation and academic grouping. Also, defined are a selected number of related terms for each construct. The purpose of these definitions is to clearly indicate important conceptual meaning that provided direction for the study. These definitions may or may not coincide with the meaning interpretations of teachers participating in the study. One of the significances of the study is to determine the congruance between prescriptive importance and teachers' own expressed conceptions of the role of grouping and evaluation in teaching.

Evaluation

Worthen and Sanders (1987) acknowledged the difficulty in achieving consensus and precision in the use of the term "evaluation:"

> ...the same terms are often used by different writers to refer to very different concepts and activities; even the term *evaluation* has been used to refer to so many disparate phenomena that the result is a confusing tangle of semantic underbrush through which the student of evaluation is forced to struggle. (emphasis in the original, p. 21)

The terms, test, measurement, assessment, and diagnosis are occasionally used interchangeably with the term evaluation. In the literature, their respective meanings are not the same, however, and should be defined more precisely at this point.

Test is the narrowest of the four terms and refers to the presentation of a standard set of questions or items to be answered. When a person is tested, the answers to such a series of questions can be used to obtain a numerical value or score of a characteristic of that person (Mehrens & Lehmann, 1984). Two major categories of educational tests are defined: norm-referenced and criterion-referenced.

When an individual's test performance is referenced or interpreted in relation to the performance of others in a defined group, the resulting score is said to be norm-referenced. For example, a child who scores in the 88th percentile in Reading on the Metropolitan Achievement Test has performed better than 88% and not as well as 12% of the children who have taken the test. Tests especially built to describe how well one has performed in relation to others in the norm group are called *norm-referenced tests* (Glaser, 1963).[5]

When an individual's test performance is referenced or interpreted in relation to a defined domain of content, skills, attitudes, or behaviors, the resulting score is said to be criterion-referenced. For example, a child who answers 24 out of 25 items correctly on a test that measures the ability to identify the correct time to the hour, half-hour, and quarter-hour has mastered this criterion. Tests especially built to describe how well one has performed in relation to a specific behavioral domain are called *criterion-referenced tests*. (Popham, 1978; Hambleton, 1982; Nitko, 1984).

Measurement is a broader concept than testing and refers to the use of any instrument or device (including testing) which allows information about behavior, objects, and events to be described in numerical form. Kerlinger (1979), representing the dominant behaviorist view, defines measurement as "the assignment of numerals to objects or events according to rules" (p. 413). Measurement can refer to both the score obtained and the process used.

Assessment is broader still than measurement and refers to the systematic collection of information. It may involve testing, measurement, observation, and any other means of acquiring valid information about an enterprise, individual, or group of individuals. What distinguishes assessment from evaluation is that the emphasis is on

the act of data gathering - i.e., the tools teachers use for these data gathering activities; whereas, in evaluation, emphasis is on the act of judging and deciding. The purpose of assessment is to better understand an area of concern. For example, why is a child having trouble summarizing and critically evaluating a lengthy reading selection, when she/he had little trouble reading the individual words to the passage correctly?

Diagnosis is a broader concept than assessment and is sometimes similar in scope to evaluation. It is a term frequently used in association with finding out about and describing the strengths and problems of individual learners in reading. It is a form of evaluation based on multiple assessments and, like evaluation, involves making judgments about the adequacy of student performance and the factors that may be helping or hindering the performance. It involves purposeful data gathering and the formation of hypotheses about ways to correct the problem(s) being diagnosed (Alexander & Heathington, 1987). A temporal delimitation distinguishes diagnosis from evaluation. Whereas evaluation is a continual process that can be invoked on practically any aspect of schooling and at a variety of levels, diagnosis is the determination of conditions that are existing at a particular time about a specific situation or problem.

An attempt has been made to carefully delimit and more precisely define several key terms often associated with the term evaluation. As a way of bringing closure to this part of the Meaning of Terms section, it is fruitful to mention at least three different broad conceptions of evaluation in education that have co-existed for the last 50 years. It was anticipated that blends of all three of these conceptualizations would emerge in the descriptive data cumulated in this study.

First, the rapid ascendancy of the so-called "Measurement Movement" in the early twentieth century created the conceptualization that measurement and evaluation

were synonymous. This orientation is still widely accepted in positivist/behaviorist approaches to classroom teaching and is typified in the writings of such measurement specialists as Sax (1980) and Hopkins and Stanley (1981).

Second, impressionistic judgment, an ancient approach to evaluation that has guided educators throughout history (Schubert, 1987), may still be the most widespread interpretation and use of evaluation at the classroom level, despite the emergence of many alternative conceptions and sophisticated techniques (see discussion in Chapter II on "considering evaluation in contemporary practice"). This orientation places faith in the professional judgments of those presumed to be experts (e.g., teachers), whether or not the criteria used in reaching those judgments are clear.

Third, a conception of evaluation emerged during Tyler's work on the Eight Year Study of the 1930s. Since that landmark study, a prevalent view of evaluation has been the process of comparing performance data with clear behavioral objectives. Other important contributions of the Eight Year Study to contemporary evaluation practice included: the use of a wide array of evaluation instruments and procedures to collect data; and, the recognition that the informal, reflective dialogue between practitioners, students, and scholars, could also be a meaningful form of evaluation (Smith and Tyler, 1942).

In short, perhaps the present sudy's most important contribution to knowledge is in providing thoughtful interpretation of the local meaning teachers give to selected aspects of student evaluation as expressed through their spoken language.

Academic Grouping

Academic grouping is a generic term that encompasses all school-based efforts to group students by ability, needs, or aspirations (Meier, et al., 1989). Two types of academic grouping are defined - ability grouping and curriculum tracking.

Public school teachers are frequently faced with the challenging problem of successfully accommodating the learning needs of 20 to 40 students in a single classroom. As observed by classroom researchers (e.g. Cuban, 1984; Goodlad, 1983), administrators and teachers have usually opted to solve the problem of pupil heterogeneity by using variations of the conventional approach of whole class instruction/recitation/seatwork. *Ability grouping* - the process of sorting students into instructional groups according to measures and/or judgments of the students' ability - was originally conceived as an alternative solution to the problem of reducing individual differences in achievement rates. Although the term "ability grouping" can refer to a variety of school and classroom grouping patterns, two general types of ability grouping identified in the literature by Slavin (1987) and Good & Brophy (1987) were pertinent to this study - between-class and within-class patterns.[6]

Between-class ability grouping refers to the process of assigning students to classes on the basis of test scores or other information about the students in order to make the class as *homogeneous* (students who are perceived to be alike on one or more selected variables) in student achievement as possible. (Good & Brophy, 1987) In his meta-review of elementary grouping practices, Slavin (1987) identifies several sub-classifications of between-class ability grouping. The arrangement where students are put in classes on the basis of their general academic ability is termed, *ability-grouped class assignment*. This process of sorting students according to indicators of their general ability is commonly called *curriculum tracking* in the United States and *streaming* in Great Britain. Oakes (1985) has observed that this form of between-class ability grouping tends to be permanent once established (i.e., "bluebirds are always bluebirds"), and that negative social and academic effects accrue to the youth grouped in

11

the lower tracks. More recent studies confirm disturbing trends in unfair treatment, inferior instruction, lower expectations, and disproportionate placements for Black children, for example, who are tracked (e.g., Race, Class and Education: The Politics of Second Generation Discrimination, co-authored by Kenneth Meier, Joseph Stuart, and Robert England, 1989).

To help offset these perennial concerns of equity and quality in educational practice as they relate to ability-grouping, many schools now employ *regrouping within grades by subject*. In this pattern, students may be assigned to *heterogeneous* (students are perceived to be unalike on selected variables) classes for part or most of the day, then regrouped for instruction in specific subjects (commonly reading or math in the upper elementary grades). Class assignments are based on measures and/or judgments about the student's ability or achievement for the specific content alone. There is modest, but not recent, evidence to support this form of ability grouping provided instruction is paced appropriately and students are not regrouped for more than one or two subjects. (e.g., Morris, 1969)

Variations of the above pattern include the Joplin Plan, where students are assigned across age and grade levels to a reading class strictly on the basis of reading achievement level. Interestingly, of all the forms of ability grouping, Slavin (1987) reports that this non-graded, multi-age grouping plan of the 1950s has shown significant positive effects on student learning compared to traditional homogeneous or heterogeneous class assignments (e.g., Floyd, 1954; Hillson et al., 1964).

Within-class ability grouping refers to the placement of students into small homogeneous groups for instruction rather than rely on whole class methods (Good & Brophy, 1987). The familiar scenario of the self-contained classroom where the teacher

has arranged students for Reading instruction into three groups of higher, medium, and lower-achieving is the most pervasive pattern of within-class ability grouping. However, even when using a between-class grouping plan, the classroom teacher may choose to further sub-divide the students into within-class ability groups. Curiously, despite the near universal use of within-class ability groups for teaching reading in the primary grades, there are no clear research data comparing the effects of this approach over others in beginning reading.

It is useful to define what is meant by "perceptions" at this point. A perception refers to information about external objects or environmental processes gained through the senses. A perception can also be described as an insight, based on observations of and subtle discriminations that lead to choice and action. In the present study, both meanings apply. Hence, teacher perceptions refer to the conscious, reflected, articulated insights of teachers.

In sum, the present study extended the descriptive research that is needed in contrasting specific teacher practices in and across ability grouping arrangements. By foregrounding the perceptions of teachers, this study suggests useful points of merger in the vast but discrepant body of research on ability grouping and classroom pupil evaluation.

Significance Of The Study

The significance of the inquiry will now be further suggested. There were three basic parts of the present study. The first was a description of criteria that teachers use to regroup students within the classroom for reading instruction. The second was a description and comparison of how teachers obtain information about the learning progress of their pupils who are in instructional reading groups within the classroom.

The third was an analysis of how selected teachers used the information obtained about the reading progress of their pupils. The significance of this study lies in the contribution it makes to helping resolve important problems in curriculum theory, practice, and research.

Curriculum Theory

One persistent problem for curriculum theory that might be better understood as a result of this study is the role of evaluation in the work of teachers. Also, implications of the research from interpetive and critical theoretical perspectives are briefly suggested.

The role of evaluation in the work of teachers. Evaluation is a complex concept with a relatively immature history in educational practice. Educators frequently misconceive the role of evaluation in curriculum and instruction. The construct of evaluation has been copiously defined by curriculum scholars, measurement specialists, ethnographers, systems analysts, and many others representing an array of disciplines in the social sciences. The concept of teacher-as-assessor or teacher-as-evaluator has not received significant attention until recently (e.g., Stiggins et al., 1985). The research of Stiggins and colleagues has focused more on the narrower concepts of classroom testing and assessment. The scope of the present study includes documentation regarding selected issues of classroom evaluation - a broader, more qualitative concept. And, while the practice of evaluating pupil learning progress has been performed and studied at many levels of education enterprise (including national, state, district, and school levels), practices of teacher-evaluators while in this function of their teaching role have not been amply studied. One of the significant aspects of this study, then, is to make a contribution to the existing literature about teacher as evaluator. Understanding the

meaning of evaluation from the teacher's perspective is a crucial focus of inquiry since it is the classroom teacher who is the prime evaluator of pupil learning.

Implications for interpretive and critical theory. Studies of ability grouping have typically focused on descriptions of the various structures of ability grouping or sought to identify various effects of the structures on student achievement, self-esteem, or self-concept. Due to the pervasiveness and stubborn persistence of the structure of ability grouping, attention in this study is turned to an interpretation of the speech teachers use to detail selected practices for pupil evaluation occurring in structures of ability grouping. This analysis is also important because it may lead to better understanding of how children are treated in various instructional groups.

Further, critical theorists often turn their attention to the study of problems associated with different aspects of the so-called "hidden" or "implicit" curriculum and how they may be affecting children's learning. Glatthorn (1987) defines the hidden curriculum as, "Those aspects of schooling, other than the intentional curriculum, that seem to produce changes in student values, perceptions, and behaviors" (p. 20). Some aspects of the hidden curriculum appear to be so intrinsic to the culture of schools that they may be impervious to change. For example, the ideology of democratic capitalism, the way in which educators define legitimate knowledge, and the dominant values of power and control represent three such constants of the hidden curriculum. Other aspects can be more readily altered by educators. Important issues related to school and classroom grouping practices - a significant alterable variable of the hidden curriculum - are raised throughout the dissertation.

Curriculum Practice

Williamson (1983) suggests that curriculum practice involves three processes: curriculum development (producing curriculum), curriculum implementation (using curriculum), and curriculum evaluation (assessing the effectiveness and worth of curriculum). There are contributions from this study to all three processes that are of value.

Curriculum development and implementation. A major and longstanding proposition in curriculum development and implementation is that the appropriate selection and clear statement of worthwhile content objectives is an essential prerequisite to sound teaching and learning. A related proposition is that the teacher's ability to collect evidence of the degree to which pupils are accomplishing the behavior and content identified in expressed objectives is enhanced in relation to the care given to the preparation of objectives for learners (Tyler, 1949). Taken together, these two propositions set the parameters of most conventional curriculum development paradigms and influence teaching paradigms as well, particularly those used in direct teaching (Charles, 1983). A study that examines the perceptions of teachers in evaluating pupil learning in ability groups may discover the attractiveness of these fundamental propositions to practitioners. If knowledge can be advanced about the meanings teachers construct for student evaluation, we may find clues for altering a cognitive structure that seems remarkably stubborn in its persistence (Goodlad, 1983; Cuban, 1984).

Further, Cuban's (1984) inquiry indicates that one of the shortcomings of curriculum research and teacher education studies has been a lack of knowledge

16

concerning what actually happens in the classroom. In a complementary way, Lieberman (1984) has suggested that:

> Teachers possess the major portion of available knowledge about teaching and learning, and it is only through a recognition of that knowledge and an articulation and understanding of it that we begin to find ways to improve schools. (p. xi)

This research builds on the premise that teachers are a primary source of information about what actually occurs in classrooms and why. The present study takes a vivid "snapshot" of teacher behavior (as interpreted through their speech) in the realm of pupil evaluation. In short, a major objective of the study was to further knowledge about a complex dimension of teaching. Through an investigation of how teachers evaluate pupil learning progress, before, during, and after group instruction, teacher educators can use resulting data to develop curricula for pre-service and in-service teachers that address existing voids in this area of competency.

Curriculum evaluation. Despite all of the varied and sophisticated methodological approaches and devices used in curriculum evaluation today, most of the responsibility for evaluating the process of public education falls on teachers as they aim at accomplishing curriculum objectives through instruction with pupils in classrooms. To the degree that knowledge can continue to be advanced about the behaviors and practices of teacher-evaluators, we will begin to more thoroughly understand the role of teachers in curriculum evaluation. Additional information about the evaluation of pupil learning progress by the classroom teacher may contribute to improving the ability of teachers to perform this important function of curriculum practice. Moreover, novice and experienced teachers may find the results useful in identifying individual priorities for evaluating pupil learning more effectively.

Curriculum Research

The potential significance of the study to alternative paradigms of research on teaching effectiveness and learning of marginal students is now considered.

Teaching effectiveness. Some investigators declare that careful teacher monitoring of student progress is one of the most important differences between effective and ineffective teaching (e.g. Kounin, 1970; Emmer, Evertson, and Anderson, 1980). Since monitoring techniques are but one cluster of the many that make up the constellation of techniques identified with the task of evaluating pupil learning, it is quite reasonable to consider that teachers found to be skilled and systematic in their approaches to evaluation will be effective teachers overall. Stated another way, students in classrooms of teachers skilled in evaluating pupil learning may attain higher levels of achievement than comparable students in classrooms of teachers unskilled or less skilled in evaluating pupil learning. This line of thinking represents the positivist orientation to research on teaching effectiveness. (For reviews of studies of discrete teaching practices and student achievement from a positivisit paradigm see, for example, Gage, 1985; and Brophy and Good, 1986).

Alternative paradigms of research on teaching effectiveness are now amply in use and are rejoined in the present study. By describing the reported practices and procedures that experienced teachers use to make judgments about their students' learning progress in Reading groups, the study attempts to make a contribution to the interpretive research paradigm on teaching effectiveness, in particular, (After Erickson, 1986) as well as to suggest implications of findings and directions for further inquiry from a more critical perspective.

Marginality and learning of minority youth. The term "marginal students" has been used by Sinclair and Ghory (1987) to refer to "students who are at the margin of attention of teachers and other school personnel as they plan and conduct the educational activities of the school" (p. vii).[7] Marginal students come from a variety of socioeconomic, racial, and ethnic backgrounds. Some of the most disturbing cases of marginality in schools have been documented for minority children.

For example, the nature and levels of literacy skills of disadvantaged black youth in the United States is unsatisfactory and unacceptable to many educators (NAEP Reading Report Card, 1985, 1990). The problem of productive school achievement is compounded for racial minorities who also happen to be poor. Study after study (e.g., CEEB, 1985) indicate that these student groups are lagging further and further behind in reading, writing, and mathematics. Studies also show that the deleterious effects of curriculum tracking and ability grouping are compounded for black and hispanic children (Goodlad, 1983; Oakes, 1985; Goodlad and Oakes, 1988; and, Meier, Stewart, and England, 1989). These significant gaps for minority youth are surfacing at the very time when pupil performance standards are being stiffened across the nation and when federal and state aid to education is declining. Once again, the American public education system is severely challenged to provide quality and equality of learning opportunity for all students. At the present time, it is unclear whether the political will and temperament to get this work done is sufficient for the task.

It should, therefore, be a national educational priority that by the next generation all black children educated by the nation's public schools and assisted by caring individuals in private and community sectors will be helped to reach advanced levels of literacy as they move into adulthood. The majority of black children in public

schools today are taught by white teachers. This demographic trend will become more acute as more minority children populate public schools and fewer minority adults seek careers as teachers. From a critical theoretical perspective, teachers need to evaluate the learning progress of minority children in culture and context-specific ways. The strategies and techniques a teacher employs to group and then evaluate the learning of an individual minority child may subtly reveal the expectations for learning that she/he holds for that child. In short, the discussion of findings of the present study may provide clues to educators about how they can be more responsive to children who currently are not being well served by the schools they attend.

Delimitations

This study has eight important delimitations. These delimitations have implications for theory and generalizability of findings.

Ability Grouping

As stated previously, ability grouping refers to the process of assigning individual learners to an instructional group based on formal and informal measures of the student's ability. In the present study, for example, a teacher might refer to a student's cumulative record for information regarding standardized test scores in Reading, previous attainments in the district-wide basal reading series, and former teacher recommendations about the pupil's strengths and weaknesses in reading. The teacher then might complement this information with her or his own inventories of the student's reading ability and performance observations during the first few weeks of the academic year. Together, these assessment data are used to make an initial decision about the placement of the student in one of several ability groups. These grouping arrangements can be structured homogeneously between-classes or in small groups within the

classroom. Hence, a learner's relative skill level, along a continuum of skills purported to be requisite to competent reading, is the primary basis for these ability grouping arrangements.

This means that other forms of grouping referred to in the classroom management literature and the literature on teaching methods - i.e., flexible grouping, grouping for more efficient administrative purposes, cooperative and competitive grouping arrangements, triad grouping and peer tutoring were not anticipated foci of the study (e.g., Cohen, 1986; Lemlech, 1988). Due to the emergent design of the study, however, the researcher did not rule out the possibility that alternative grouping strategies might be uncovered.

Evaluating Pupil Learning

Evaluation of pupil learning as it is performed by teachers refers to the process of gathering evidence, forming impressions, and placing a value on how well and to what degree a student is accomplishing the targets of the curriculum (e.g., Taba, 1962; Bruner, 1966; Jones, 1977 and many others). From an empiricist view, evaluation may mean that the teacher observes a student's performance, measures it in some way, compares it with a standard, and judges the comparison as favorable or unfavorable (Rinne, 1984). Informal discussions with teachers reveal that they may hold similar conceptions of the construct of evaluation. Many teachers view evaluation of pupil learning as making some kind of decision about whether and to what degree the students have mastered the subject matter they have been assigned to teach. *Evaluation of learning, student accountability,* and *pupil assessment* refer essentially to the same activity when referred to by practitioners.[8] To reduce confusing jargon, the term

"evaluation of learning" will be limited to the data gathering and judgment activities of teachers.

Subject Matter

A researcher investigating the ways that teachers evaluate pupil learning in ability groups might anticipate that the assessment strategies of teachers vary in kind, in frequency, in rigor, and in technical validity depending on the subject matter they are teaching. A third important delimitation of the present study is that Reading is the only subject matter domain of instruction selected for intensive inquiry.

Of all the subjects learned in school, perhaps Reading is the most essential. Reading is notorious as the subject where students are sorted and grouped according to ability. Moreover, reading is the basis for other learning since all academic subjects are reading subjects. This fundamental premise has been affirmed again and again by American educators for almost a century. In 1895, at the annual meeting of the National Education Association in St. Paul, Minnesota, members in attendance expressed through their journal proceedings and addresses:

> ...Reading and writing are not so much ends in themselves as means for the acquirement of all other human learning. This consideration alone would be sufficient to justify their actual place in the work of the elementary school. (NEA, 1895).

In 1986, in his national report on elementary education, William J. Bennett, U.S. Secretary of Education, reaffirms this consensus of educators:

> The elementary school must assume as its sublime and most solemn responsibility the task of teaching every child in it to read. Any school that does not accomplish this has failed. (p. 21)

Moreover, better readers are better writers according to Archie LaPointe, Director of the National Assessment of Educational Progress (NAEP). Unfortunately,

too many of our children exiting elementary school can't read and an even greater number can't write - or not nearly well enough. To make the case statistically, following are some selected data in Reading from 1985 NAEP reports:

- Ninety-four percent of 9 year-olds (in the third or fourth grades) could read at a rudimentary level. This means, though, that every year, after three or four years of instruction, about 200,000 boys and girls still cannot read.

- Sixty percent of 13 year-olds, (in seventh or eighth grades) could read and understand their textbook with ease. This means that, even after a complete elementary school experience, 40% of them could not. Stated another way, among 13-year-olds, only three out of five were reading at the skill level appropriate for their age, defined by NAEP as the 'ability to search for specific information, interrelate ideas, and make generalizations.' Virtually all of them possessed rudimentary reading skills. But forty 13 year-olds out of a hundred, 2 out of 5, lacked the 'intermediate' reading skills that would enable them satisfactorily to handle the books and lessons that a seventh or eight grade teacher should be able to assign. Worse, most of the minority 13 year-olds who were sampled (65% Black, 61% Hispanic) were reading below the intermediate level. (NAEP, 1985).

Experts agree that these data do not necessarily reflect a decline in the reading performance of our youth compared to recent years; rather, they indicate that many (too many) of our children and youth are not reading to their potential (Carbo, 1987). Thus although the current research centers only on reading, national trends in student educational achievement suggest indeed that it is a crucial subject for consideration.

Contemporary Teaching Practices

An important focus of the study is on present times and realms of practice as they are occurring day-to-day in elementary public school classrooms. This delimitation excluded an historical approach of how pupil evaluation in ability groups for Reading has been carried out in different bygone eras. This also means that approaches to pupil evaluation used by a teacher earlier on or in years past in his/her teaching career are

not the focus of this study. Rather, it is on what teachers say they are doing now, not on what teachers have done or may do in the future as their professional practices continue to evolve. It also excludes an investigation of what the evaluation of pupil learning in ability groups may mean to those who conduct research and make administrative or legislative decisions pertaining to school curriculum.

By recognizing the target population for the study as public school elementary teachers in selected classrooms in Western Massachusetts, contemporary evaluation practices of these teachers as a delimitation restricts the degree of generalization that may be inferred from the results of the study to the realm of those teachers and classrooms considered. Hence, the study does not purport to address other levels of schooling (e.g., secondary, post-secondary); or, deal with a comparative analysis involving the pupil evaluation practices of teachers for Reading ability groups as they are carried on in other regions of Massachusetts or the United States.

Explicit Teaching

The question of what kind of teaching by what kind of teachers requires some deliberate attention and is discussed as a fifth delimitation for the present study. The kind of teaching that is examined is classroom teaching - as it goes on in persistent fashion in public elementary schools in the United Sates where teaching has been systematically observed. The stubborn continuity of the character of instruction in our public schools has been well documented (Goodlad, 1983; Cuban, 1984). Despite the many approaches to teaching that have been introduced during recent decades, "teacher-centered instruction" where - teacher talk exceeds student talk during instruction, the use of class time is determined by the teacher (Cuban, 1984), and the range of pedagogical methods observed is exceedingly narrow, particularly when considered

alongside the diverse ways that humans learn (Goodlad, 1983) - predominates in our nation's classrooms. Although there is ample opportunity for the classroom teacher to depart from this conventional approach to teaching during small group instruction for diverse learners, there is little evidence to suggest that such departures are frequent and sustained (NAEP, 1985; Good & Brophy, 1987). Larry Cuban (1984) succinctly defines the conventional approach to teaching:

> Classroom teaching does not mean lecturing exclusively, or discussing, exclusively, or tutoring, exclusively. Rather, it means the combination of all these ways of teaching plus the classroom recitation, or relatively rapid-fire teacher questioning and pupil responding, and so-called seatwork. In addition, classroom teaching includes a variety of managerial activities that keep the whole process moving along in an orderly way. (p. 3)

Similarly, the conventional basal approach to small group Reading instruction typically begins with a review of the new or introduced vocabulary words from the flip chart accompanying the basal series; next, the teacher asks divergent-level questions pertaining to the story selection for the day; next, she asks the group to read a portion of the story silently to themselves to become familiar with the context; she then appoints or calls on volunteers to begin reading aloud passages from the story, periodically stopping to ask comprehension level and prediction questions. To close the lesson, the teacher might ask evaluative questions of the children about the story, and then assign a follow-up worksheet that tests a discrete reading skill.

Curiously, this basic instructional approach does not differ appreciably across higher, mid, and lower achieving reading groups (Carbo, 1987). And although when surveyed, teachers claim to use a wide variety of approaches to teach reading (LaPointe,

1986), it appears that the variable that is more typically altered by teachers for the various ability groups is pace rather than method (Good & Brophy, 1987).

Perhaps because of its prevalence and persistence, it is this conventional view of teaching that was deemed more pragmatic to study. Stated conversely, this means that such instructional approaches as Programmed Instruction, Computer Assisted Instruction (CAI), Bloom's Mastery Approach, The Keller Plan, Individually Prescribed Instruction, The Program for Learning According to Needs (PLAN), The Direct Instructional System for Teaching Arithmetic and Reading (DISTAR) were not anticipated to be widely used and were excluded as foci of the present study. Also, special pull-out programs that are frequently devoted to small group remedial instruction in Reading such as Chapter I services, resource room tutoring, and the use of aides or parent volunteers were backgrounded in the analysis of the present study.

Teacher Population

It was assumed that teachers more than any other group, (e.g., parents, students, administrators) would be able to provide valid data about their assessment and evaluation practices, specifically, as they apply to the individual progress students are making in learning to read in ability groups. The teachers studied were experienced public school teachers. Public school teachers have been commissioned by society to lead the enormous effort of educating the nation's children. Teaching is the central process, the heart, of education. Hence, it is public school teachers that have the major responsibility for producing desired educational outcomes in learners. If public school teachers are held in close account for the learning achievements of their pupils, then it may be reasonable to assume that this population of professionals, if experienced, has given considerable thought to 1) how they assess pupil learning progress in ability groups

for Reading; and, 2) how they use the resulting data to improve teaching and learning for students in the various groups. The study focused on teachers with three or more years of teaching experience rather than novice teachers who are still getting acquainted with the many demands of the profession.

Another reason for delimiting the unit of analysis to public school teachers is that the vast majority of youth who are in school in America are in public schools. During the 1986-87 school year, for example, 31 million boys and girls were being taught by roughly 1.45 million public school teachers in more than 75,000 elementary schools across the United States.[9] This represents by far the greatest distribution of schools, children, and teachers in the enterprise that is American elementary education.

Level of Schooling

A seventh delimitation, centers on the level of schooling selected for consideration. First, the philosophical eminence of elementary education in the lives of children is clear. William Bennett (1986), frames the case:

> Elementary education is an enterprise of vast proportions in this nation; and for each child it is an experience of unsurpassed importance. After the family, elementary school is the most influential institution in children's lives: helping to shape first and lasting views of themselves, molding aspirations and skills, and introducing them to their culture, to the universe itself. (p. 1).

Second, during the primary years (K, 1, 2), intensive activity occurs in sorting children into ability groups, particularly for Reading. During these early years, letter grades do not play as significant a role in pupil assessment as do teacher's academic and non-academic ratings about the child's progress. Nonetheless, teachers in these grades labor under an unspoken pressure to have all their pupils reading by the end of their tenure in first or second grade.

Third grade is the time in many public schools when children's learning progress is first reported in the form of letter grades for the content areas comprising elementary education - i.e., Reading, English, Writing, Spelling, Mathematics, Science, Social Studies, Health, Art, Music, and Physical Education. It is no longer a case of largely informal, mentally-stored evaluations accompanied by brief quarterly narratives; now, teachers must include summative measures of their pupils' learning and transform these and their formative evaluations into a grade or symbol of achievement for the marking period that is meaningful to the student and the student's guardian.

Similarly, the intermediate grades 4-6, represent a crucial interval in the elementary education of children. During these years, the critical appraisal of pupil achievement intensifies. Sorting practices continue and begin to stabilize during this time. At the same time, the groundwork for much of the functional knowledge that will be used throughout adult living is cultivated. That is, the seeds for basic concepts and facts, communication and computation skills, persistent cultural values, and attitudes and appreciations necessary for getting along with others are sown. Moreover, as children's school career progresses they are expected to use Reading as an integral tool for learning literature, science, mathematics, social studies, and the arts and humanities. It is in grades 5 and 6 that they begin to use reading for this purpose in a more serious and sustained way than ever before in their elementary education. It is quite telling, therefore, to find out how reading is perceived to be assessed by teachers in the upper elementary grades.

Third, grades 5 and 6 often represent the last chance for direct reading instruction for public school learners. Due, in part, to the crystallizing effects of ability grouping for lower achieving students, those who are not reading very well are rather

easy to distinguish by fifth and sixth grade (Slavin, 1987). By focusing the inquiry in these two grades, it is presumed that useful knowledge can be produced about how the progress of children grouped as low, medium, and high achievers in Reading are being evaluated and then attended to by teachers.

Teacher Perceptions

Eichelberger (1989) suggests that perceiving is not an automatic process by which whatever exists in the world is translated directly into our minds. (p. 11) Rather, our experiences, ideologies, values, religion all affect how we perceive. Perceptions can also be related to the needs of the individual. Hence, one of the disadvantages of using perceptions as a basis of knowledge about reality is that because perceptions often reflect personal needs, they may distort reality. People's perceptions can be further stymied by habitual ways of thinking and behaving.

Moreover, there may be a difference between what people say they do and what they actually do. Data furnished by self-report, even when sensitive efforts have been made to put respondents at ease, have potential distortions that are difficult to control. Respondents may have an inclination to exaggerate, resist, or vent a variety of other hidden agendas. Many professional and lay people have an aversion to questionnaires and are uncomfortable in interview situations. For these reasons, inferences or conclusions made on the basis of perceptual data must be considered with these delimitations in mind.

The following chapters constitute a detailed description of the present investigation. Chapter II provides a conceptual background for interpreting the meaning and use of evaluation, establishes the theoretical orientation of the study, and presents a selected review of literature. Chapter III describes the design and research procedures

of the present study. Chapter IV reports the analysis of data and interprets the

research findings as they relate to the three major research questions. Finally, Chapter

V summarizes the investigation, provides a discussion of implications of findings, and

suggests possible directions for educational practice. The chapter concludes the study by

recommending further research.

CHAPTER II

THEORETICAL ISSUES AND REVIEW OF RELATED LITERATURE

This chapter consists of five main sections. First, the review concentrates on the selected approaches to curriculum evaluation culled from the literature from 1965 to the present. Once a context for classroom pupil evaluation has been established, the theoretical perspective underpinning the inquiry is set forth. Next, it is possible to review more purposefully three selected literatures which support the importance of the three research questions being investigated in the present study. The third section reviews exemplary studies of ability grouping for Reading instruction to suggest the importance of discovering the criteria teachers use to group students for instruction. Fourth, selected studies on classroom pupil evaluation are analyzed for their role in providing a better understanding of this complex function of teaching. Fifth, selected studies from an emerging branch of inquiry on teacher decision-making are described to provide support for investigating the substance of Research Question 3 - how teachers use data obtained from classroom pupil evaluation.

Considering Evaluation in Contemporary Educational Practice

In twenty short years, following the enactment of the 1965 Elementary and Secondary Education Act, educational evaluation had become a full-blown field of inquiry and practice. Able scholars, advances in measurement technology, and the demand for accountability in a $120 billion a year industry (education) hastened the development of educational evaluation. By the start of the 1980s, federal expenditures for evaluation were over 300 million dollars. Financial support for program evaluation had increased dramatically at state and local levels of government as well. Widespread

demand for expertise in evaluation prompted many educational researchers to move into the field full time.

In more recent years, political consensus seems less supportive of the democratic principles of equity and social justice for all. Administrative policies at the federal level have adversely affected educational and social programs for disadvantaged adults, youth, and families. Meanwhile, state governments plunge deeper in debt, many mired in unprecedented fiscal crises. The results are deep cuts in budget outlays for education at all levels. This condition, among others, may partially explain the decline of educational evaluation as practiced and conceptualized in the late 1960s, 1970s and the first half of the 1980s.

Yet, the need for evaluation has not diminished. Teachers feel the press from principals, superintendents, parents, and school committees of having their pupils achieve to desirable norms on standardized tests. They feel their own conscience tugging at them to determine better ways to evaluate successful and less successful learners. But what of the meaning of evaluation? Have we taken sufficient time to explore the many dimensions of this important concept? Is contemporary practice adequate for the complex enterprise that is education? Has the value of formative assessment and common-sense information gathering strategies at the classroom level been overlooked in our modern fetish to apply formal designs and sophisticated statistical techniques to evaluation conducted at the program level? Have the meaning-perspectives of teachers toward evaluation been sufficiently investigated to understand the role of evaluation at the classroom level?

This section of the literature review addresses these questions by organizing discussion into three parts: the first part describes where the meaning of a concept

comes from; the second part gleans important concepts, assumptions, and methodology that so-called evaluation models can contribute to the professional practice of teachers; the third part delimits the term evaluation from a study of educational programs to an integral dimension of teaching.

Scholars' Frame of Reference

One place to look for the meaning of a concept is in the frames of reference expressed by scholars. The frame of reference is the "lens" through which one perceives the experiential world. This lens contains many sources of error, for it is colored by such variables as one's overall view of the world, values, attitudes, personal history, biases, educational training, and occupational roles. Nonetheless, the frame of reference structures the rules for reducing ineffable pure fact to described fact - in short, the rules of concept formation - and gives rise to the major concepts, propositions, and theories that are used to explain reality (Meehan, 1969).

Scholars understand the importance of this epistemological network (only briefly introduced here) and so are deliberate in structuring the frame of reference upon which rest the concepts, propositions, models, and theories in their discipline or field of inquiry. The transmission of concepts can be traced through mass as well as esoteric channels within a culture. Scholars, through their discourse, lend definition to a particular concept by helping to clarify what the concept is, what it consists of, what its purposes and criteria are, what the important elements and dimensions are, and how these are related to each other. Practitioners, by virtue of what they actually do, make a concept meaningful. All of us, through membership in a democratic society which encourages intellectual rigor and creative thought, are invited to participate in the

"language game" (Hirst, 1974) that continually shapes and recasts the meaning of a concept.

Hence, the constructs that bind the field of educational evaluation together have been defined by measurement specialists, curriculum specialists, ethnographers, linguists, systems analysts, and many others representing a broad array of disciplines including psychology, sociology, anthropology, political science, philosophy, economics, law, arts and humanities, and mathematics. The meaning of the concept of evaluation, however, has largely been understood in relation to the practices and purposes to which it has been put. In education, the view of evaluation that is most prevalent is that of an evaluator commissioned by decision-makers or sponsors to evaluate an educational program. Data collected and interpreted by the evaluator are for the purpose of judging the worth, quality or merit of the program and for assisting in decisions of whether to improve, maintain, or terminate the program. Within this broad conceptualization, there are many different approaches (sometimes referred to as models) and subsequently many different strategies, methods and techniques used to direct an evaluation. These vary from evaluator to evaluator and from situation to situation.

Scholars would agree perhaps that an important inference to make is that the evaluator should strive to employ those methods which are best suited for producing the desired information. Arguments about the intrinsic and universal superiority of one method over another are not fruitful. Instead, efforts could be focused on helping teachers, for example, attack problems of evaluation (of pupil learning and instruction) with the widest array of conceptual and methodological tools possible and that such problems demand (Trow, 1970). The next part extends the discussion by describing

concepts and methodologies from program evaluation that may be useful to the teacher-evaluator.

Contributions from Program Evaluation

The following review points out some of the important distinctions between selected evaluation approaches or models used to evaluate programs at the national, state, district, and school levels. It is recognized that these descriptions over-simplify the evaluation approaches actually used. Experienced evaluators try to blend or adapt methodologies to match the specific purposes of evaluation. Moreover, these models are complicated by the assumptions, ethics, epistemology, and politics underlying their use (Sjouberg, 1975; Srouff, 1977; House, 1978; Talmadge, 1982). The intent here is synthesis for the purpose of helping the teacher-evaluator to infer meaning and to cull useful concepts from contemporary practice in educational evaluation for creative application with learners, rather than detailed analysis for the purpose of critiquing and classifying existing evaluation paradigms and methodologies.

Worthen and Sanders (1973), Popham (1975), Stake (1976), House (1978) and others have developed useful taxonomies for the major models used to conduct evaluations of educational programs. Eight models are described in turn - Systems analysis, Behavioral objectives, Decision-making, Goal free, Art criticism, Accreditation, Adversary, and Transaction. Each is described in terms of the purpose, audiences, methods, outcomes, and typical questions or concerns espoused for it by its leading proponents.

Systems analysis. A model that has predominated in the evaluation of many federal and state social action programs is systems analysis. Systems analysis is concerned with analyzing the program "system" into its component parts to determine

which parts (inputs) are interacting in what ways to produce the effects (outputs) of the program. Important variables and outcomes are specified in logical and deductive fashion in advance, quantified, and where possible, manipulated experimentally to test hypothesized relationships. As Michael Patton explains:

> Systems analysis requires (1) identifying the important input and output variables for programs, (2) quantifying those input and output measures, and (3) statistically analyzing the relationships between program inputs and program outcomes. Inputs are quantities such as program budget, staff size, staff-client ratios, client characteristics, baseline performance levels of clients (pre-client performance levels after the program (post test scores)), placement rates, monetary value of new client skills, new client wage levels or equivalents, and so on. (Patton, 1980, p. 50).

Methods of investigation used in systems analysis include planned programming budget systems (PPBS), linear programming, planned variation, and cost-benefit analysis. Aimed at providing information for managers and economists typical questions for the evaluation are: Are the expected effects being achieved? Can the effects be achieved more economically? What are the most efficient programs? What variables cause or contribute most to the efficient and productive operation of programs?

Behavioral objectives. The behavioral objectives approach has dominated educational evaluation since its inception. Specific domains of behavior stated in the form of written objectives, are sampled, measured against student performance, and assigned a score. The technical quality of tests used and the validity of data collected are important standards of success for the evaluation. Evaluation consists of those methods used to compare student performance with behaviorally stated objectives. Ralph Tyler (1949) sets forth the rationale for this approach succinctly:

> The process of evaluation is essentially the process of determining to what extent the educational objectives are actually being realized by the program of curriculum and instruction. However, since educational objectives are essentially changes in human behavior, that is, the

36

objectives aimed at are to produce certain desirable changes in the behavior patterns of the student, then evaluation is the process for determining the degree to which these changes are actually taking place. (pp. 105-106)

Although educational outcomes specified in terms of specific student performance criteria have been largely measured by standardized achievement tests and statistically analyzed to compare individual and groups of students, Tyler's recommendations concerning methods of appraisal were more inclusive. Not only pre and post measures, but still another point of evaluation, made sometime after instruction had been completed, was necessary in order to measure the sustained effects or permanence of learning.

Tyler (1949) further suggested that:

Since evaluation involves getting evidence about behavior changes in students, any valid evidence about behaviors that are desired as educational objectives provides an appropriate method of evaluation. (p. 107)

This implies that not only paper and pencil tests, but for example, observation, interviews, questionnaires and self-reports, products, and records of students all could be useful means for getting evidence about behavior changes occurring in learners.

Intended for use by program managers and measurement specialists, the behavioral objectives evaluator asks: Are the students achieving the expressed objectives? Is the teacher producing the desired behavior changes in students?

In addition to Tyler, W. James Popham (1975) is a leading contemporary advocate of behavioral objectives and the refinement of criterion-referenced testing as a device to measure their attainment. Major precursors of the behavioral objectives movement might be traced from the work of William James (1890) and Edward L.

Thorndike (1906) in psychology, and Franklin Bobbitt (1918; 1924), W. W. Charters (1922), and Harold Rugg (1927) in curriculum.

Decision-making. In this approach, the evaluation is structured by the decisions to be made. The evaluator is to supply information relevant to these particular decisions. Decisions-makers refer primarily to administrators and managers of educational programs, although the term has broadened to include other decision-making groups sponsoring the program, i.e. legislatures, school committees, executive boards and councils, etc. The primary purpose of evaluation is to provide information for decision-making.

Daniel L. Stufflebeam and the Phi Delta Kappa National Committee on Evaluation (1971) designed the CIPP model of evaluation as an aid to rational decision-making in the day-to-day management of programs. Developed in part in response to the demand for accountability of newly operating programs that had emerged from the decade of the sixties, the CIPP model combines concepts and procedures of the systems analysis and behavioral objectives models. The model is characterized by four stages or types of evaluation emphasizing costs, efficiency, and options for decision-makers. The four stages are: 1) Context evaluation - information used to develop rationale and to determine project objectives; 2) Input evaluation - information used to determine how to utilize resources to achieve objectives; 3) Process evaluation - information used to detect problems in the procedural design and implementation of the program; 4) Product evaluation - information used to interpret attainment or outcomes. Formative (during the course of) and summative (end) measures are used to determine these outcomes. Product evaluation information is also used for recycling decisions through the context, input, and process stages.

Stufflebeam (1971) proposes a definition of evaluation which complements the CIPP model:

> Evaluation is the process of delineating, obtaining, and applying descriptive and judgmental information for decision-making and accountability.

Meanwhile in the field of educational evaluation, Thomas Hutchinson and a group of colleagues at the University of Massachusetts at Amherst (1973) were working to develop a fully explicated set of rules and procedures by which to conduct evaluations for decision-making. Depending upon one's "frame of reference," the Decision-Maker Oriented Evaluation Methodology developed by Tom Hutchinson et al. is a practical approach to evaluation. The authors define their methodology as a "logically deductive and empirical process for arriving at a complete, fully operational set of rules and procedures for accomplishing a definable purpose" (p. 1). The purpose of this evaluation methodology is to provide data about an enterprise for decision-making.

Further, the authors provide operational definitions of the key terms contained in the overall definition and purpose of the methodology. An "enterprise" is any identifiable process or thing. "Data" are information provided for decision-making.

Three criteria are given as standards to which the methodology can be held - efficiency, focus, and completeness. These criteria are operationally defined as follows:

1. The data provided are used by a decision-maker (efficiency).

2. The data provided are focused on the more important decisions rather than the less important ones (focus).

3. The data provided are as complete as possible given resource limitations (completeness). (p. 4).

The evaluation methodology is completely successful when "efficiency" is at 100%, focus is at 100%, and completeness is at the maximum given the resources available.

Though derived from a strong empiricist, "objectivist epistemology" Patton (1980) suggests that:

> The decision-making model does not imply any particular methodological stance. Indeed, it is the most open of all the models to a full variety of methodological strategies, both pure and mixed forms. The methods to be used depend on what evaluation information is needed to help make specified decisions. (p. 58).

In the Stufflebeam model, key questions are: Is the program effective? Which parts are contributing to the success or failure of the program? In the Hutchinson methodology the key question is: What are the decision questions for which information is needed?

Goal-free. An approach to evaluation that emphasizes program effect has been introduced by philosopher-evaluator, Michael Scriven. "Goal-free" evaluation, as Scriven calls it, is uninterested in what proponents of the program have to say about it or what its explicit goals are. Though Scriven embraces the value of goals for planning purposes by program developers, he does not believe that a program evaluator needs to be made aware of these goals in order to carry out his primary role - to judge the worth of the program. Hence, the purpose of Goal-free evaluation is to determine the merits and effects of the program, intended or not; to evaluate the "treatment" without knowledge of what the "treatment" is supposed to do.

To avoid being influenced by "contaminated," i.e. subjective information, Scriven's evaluator relies on a highly structured checklist of 13 indices or evidences as the means for trying to assess all the effects of the program. Scriven contends that the goal-free evaluator's instruments (e.g. needs assessments, observation scales, constructed tests,

weighted scales, etc.) will reveal, in objective fashion, the actual effects of the program. The effects are then matched against the goals to establish and justify the worth of the program. Scriven explains the need for Goal-free evaluation strategies to complement Goal-based approaches:

> ...I became increasingly uneasy about the separation of goals and side effects. After all, we weren't there to evaluate goals as such - that would be an important part of an evaluation of a *proposal*, but not (I began to think) of a product. All that should be concerning us, surely, was determining exactly what effects this product had (or most likely had), and evaluating those, whether or not they were intended.

> ...Furthermore, the whole language of "side effect" or "secondary effect" or even "unanticipated effect" ...tended to be a put-down of what might well be the crucial achievement...Worse, it tended to make one look less hard for such effects in the data and to demand less evidence about them, which is extremely unsatisfactory with respect to the many potentially very harmful side effects that have turned up over the years.

> It seemed to me, in short, that consideration and evaluation of goals was an unnecessary but also possibly contaminating step. I began to work on an alternative approach - simply, the evaluation of actual effects against (typically) a profile of demonstrated needs in this region of education. I call this goal-free evaluation (GFE). (Scriven, 1971, pp. 1-2) (italics in the original)[10]

In sum, the goal-free evaluation requires the evaluator to suspend judgment about what it is the program is trying to do and focus instead on finding out what it is that actually happens (as a result of the program). The essential questions for the goal-free evaluator are: What are all the effects of the program? What are the merits of the program?

Art criticism. Evolving from the traditions of art and literary criticism is the model of an educational critic, one who has, through experience and training, perfected skills in judging the important facets of educational programs (House, 1978). Elliot Eisner (1975) has developed two concepts, educational connoisseurship and educational

criticism, to enrich and broaden the repertoire of intellectual structures we might use to understand school and classroom life. Eisner believes that informed educational criticism may give a teacher a view of his or her own teaching that he simply otherwise would not possess. As Eisner explains:

> ...Because I believe teaching in classrooms is ideographic in character, that is, I believe the features of classroom life are not likely to be explained or controlled by behavioral laws, I conceive the major contribution of evaluation as contributing to a heightened awareness of the qualities of that life so that teachers and students can become more intelligent with it. Connoisseurship plays an important role towards this end by refining the levels of apprehension of the qualities that pervade classrooms. (p. 8).

Connoisseurship, then, is the art of appreciation. And criticism, the art of disclosure. Criticism, as Dewey has noted in <u>Art as Experience</u> (1934), has at its end the re-education of perception. Eisner elaborates:

> What the critic strives for is to articulate or render those ineffable qualities constituting art in a language that makes them vivid. In doing this something of a paradox exists. How is it that what is ineffable can be articulated? How do words express what words can never express? The task of the critic is to adumbrate, suggest, imply, connote, render, rather than attempt to translate. In this task, metaphor and analogy, suggestion and implication are major tools. The language of criticism, indeed, its success as criticism, is measured by the brightness of its illumination. The task of the critic is to help us to see. (p. 8).

What is it that the educational critic does when he writes criticism of a classroom, or a set of curriculum materials, or a school? Eisner suggests three things. The critic describes, interprets, and evaluates or appraises what he sees. These distinctions are not intended to be independent or sequential.

> ...The descriptive aspect of educational criticism is an effort to characterize or render the pervasive and sheerly descriptive aspects of the phenomena one attends to. For example, critical description might tell the reader about the number and type of questions that were raised in

42

class, the amount of time that was spent in discussion, or the kind of image or impression the teacher or the room gives to visitors...

...The interpretive aspect of educational criticism represents an effort to understand the meaning and significance that various forms of action have for those in a social setting. For example, just what do the extrinsic rewards for reading mean to the third graders who keep charts of the number of books that they have read?...What kinds of messages are being given to students by the allocation of time and its location in the school day to the various subject matters that constitute the curriculum?...

...The third aspect of criticism is evaluative. It asks, 'What is the educational import or value of what is going on?' To deal with educational import of classroom life...is to make some value judgments about it with respect to its educational significance...The critic uses what he or she sees and interprets in order to arrive at some conclusions about the character of educational practice and to its improvement. (p. 9)

Finally, the questions that are important to the educational critic are many. These are often focused on the discourse within a particular classroom. Eisner frames possible questions for critical attention, for consideration, and in so doing, captures the essence of this approach to evaluation.

> ...How do the children participate? What is the quality of what they and the teacher have to say? To what extent do they participate both psychologically and verbally in what transpires? Is their enthusiasm feigned or real? Is what they are learning worth their time and effort? And just what are they learning? Is it what is being taught, or are they learning other things that are carried by the manner of teaching and the organization and structure of the school day? What about the materials they use, the textbooks, the learning kits, the visuals with which they come in contact? What do these materials teach? How are they laid out? What does their format convey? What messages are held between the lines of textbooks which for so many children occupy central roles in their school experiences?

> ...What about the relationships among the children themselves? Is it competitive or cooperative? Is the class a collection of individuals or a community? What is the pervasive quality of educational life that children in this particular classroom lead? How is time allocated within the school day? How are the various subjects taught? What values are conveyed by the ways in which time and space decisions have been made?

...What is the quality of work that children create? What is the character of their expression, verbal, written, visual, musical? Over time, what kind of development is evident? In what ways is the development of intellectual curiosity and autonomy displayed? In what ways are they treated when they are expressed? (pp. 14-15).

Eisner, a prominent art educator, has been working to refine the notion of educational evaluator as connoisseur and critic to the design and evaluation of school programs (See Eisner, 1985).[11]

Accreditation. For nearly sixty years, educational institutions have cooperatively united to evaluate each others' programs. This usually is done by a team of experienced (outside) professionals visiting on-site. The team relies heavily on analysis of program documents, informal interviews, and site visit observations as well as their own knowledge and expertise. Meanwhile, the local professionals have collected and prepared information about their program according to a set of external criteria. The accreditation team reviews, then commends or disapproves of the local programs.

For example, typical observation methods of accrediting teams reviewing teacher preparation programs in higher education include: checking student files, transcripts, and pre-practicum and practicum report forms; holding conversations with students, faculty, and cooperating school personnel to determine whether candidates for teacher certification meet pre-requisite, semester hour, and practicum requirements for the standards of the effective teacher; reporting their judgments on these data via a formal written report (Quist et al., 1985).

A related and frequently used approach to evaluation at the institutional level is the self-study. An institutional self-study is generally undertaken by a committee of faculty and administrators whose task is to review and make recommendations for increasing staff and program effectiveness. The self-study is often called into service as

44

a result of internal pressures - e.g. needed curriculum revision, staff imbalance, declining enrollments, high attrition rates among sub-populations of students, reduced funding, etc. - without any external mandate. As a result, self-studies have the commendable value of keeping problem-solving responsibility at the site of the problem (Prendergast, 1983).

Another related approach to evaluation at the institutional and societal levels, as well as the program level, is the panel of leading citizens - people who are held in high esteem, who have a strong sense of social responsibility, and who are respected for outstanding achievement of some kind.[12] They may follow their intuition, professional judgments, or be guided by an experienced chairperson. The outcome of the work of this group is typically a formal report recommending directions and priorities for improvement.

Adversary. As concern about the politics and pressures of evaluation grew in the late sixties and early seventies, and as critics argued with increased fervor that single evaluators could not maintain impartiality and objectivity throughout the evaluation process, support for the conceptual model of the advocate-adversary grew. Adversary evaluation is a model of evaluation derived from procedures used in jury trials and administrative hearings in the field of law. The purpose of the model is to present the pros and cons of the issues that are relevant to the groups of people or stakeholders involved in the program. To accomplish this purpose, adversary evaluation relies primarily on human testimony and the debate skills of the members of the adversarial teams as the methods used to collect and present information about the program.

One of the architects, Robert L.Wolf (1975), frames the context and intent of the adversary model:

> The forum for carrying out such procedures is what I am calling an
> educational hearing. The hearing is not intended to totally replace

45

existing designs for the collection and analysis of evaluation evidence, but rather to provide a more effective way of seeking and presenting balanced factual data. Currently, many assumptions, rationales, methods of data collection, and analysis of evaluation reports are allowed to pass unchallenged. The judicial approach provides for the structured consideration of alternative arguments and inferences to keep the evaluation fair and honest. Unlike true adversary proceedings in the law, where the adversaries object is to win a case, educational hearings are aimed at producing broad program understanding, exploring the complexity of educational issues, and keeping at least two sides of the truth alive. (pp. 185-186)

Wolf and others have described the process of adversary evaluation. The approach has four basic stages. The first stage is to generate a broad range of issues. To do this, the evaluation team surveys the major stakeholders in the program - e.g. users, managers, sponsors, community supporters, etc. - to determine what they believe are the relevant issues.

The second stage involves reducing the list of issues to a manageable number. One method for doing this is to have a sample group of respondents list the issues in order of importance.

The third stage is to form two opposing evaluation teams (the adversaries) and provide them the opportunity to prepare arguments for or against the program on each issue. As part of their methodology, teams may interview witnesses, study existing evaluation reports, and collect new or additional data.

The final stage is to conduct pre-hearing sessions and a formal hearing. In the formal hearing the opposing teams present their cases and supporting evidence before the jury, the program's decision-makers, for judgment.

The question of most concern to these decision-maker jurors is: What are the reasoned arguments for and against the program? Owens (1973), Kourilsky (1974),

Levine (1974), and Wolf (1975) have described case studies using the adversary model in evaluating educational programs.

Transaction. This approach concentrates on describing educational programs as perceived by the people most closely involved with them. Based on these varied perceptions, the evaluator makes an interpretation of the program's worth as seen through the eyes of children and adults who have a stake in it. The evaluator adopting this approach uses subjective methods of inquiry, including observation, interview, questionnaire, narrative and case study to investigate the major components of the educational process - the school, the program, the classroom.

The transactional model is derived from transactional psychology which makes certain presuppositions about perception. Chief among these presuppositions is that the perceiver cannot be separated from the context in which his perceptions occur. She (the perceiver) affects and is affected by the situation, and thereby, is part of the transaction. (Ittleson and Cantril, 1954). Hence, to accomplish the principal aim of this model - i.e. a comprehensive portrayal of the program, with fidelity to the many important perceptions and expectations of it - means to study the particular transactions in which the perceptions can be observed.

Robert E. Stake (1972), a leading proponent of this approach eloquently states the case for portrayal of the whole vs. analysis of a piece:

> The whole cloth of an educational program is a grand accumulation of intents, transactions, and outcomes. The teachers intend to deliver on many promises and to take advantage of many targets of opportunity. Students and parents have their expectations and apprehensions. Community leaders, social critics, and educationists have "viewed with alarm" and "pointed with pride." Each child brings his own complex of convictions, misunderstandings, and propensities and takes away some of those and still others. Each classroom is a community, with rules and stresses and competition and compassion. Yesterday's subgroups are not tomorrow's. Things are learned, unlearned much as shoelaces are

knotted, untied, broken, and retied. An educational program has
countless objectives, many of them dormant until a crises arises. The
priorities vary over time from person to person. No statement of
program objectives ever devised has come close to representing the real-
word intents of the people involved in an educational program. (p. 2)

In another published piece, Stake (1974) describes in concrete and practical
terms how to actually do a "responsive evaluation," an example of the transactional
approach to evaluation.

> To do a responsive evaluation, the evaluator conceives of a plan of
> observations and negotiations. He arranges for various persons to
> observe the program, and with their help prepares brief narratives,
> portrayals, product displays, graphs, etc. He finds out what is of value to
> his audiences, and gathers expressions of worth from various individuals
> whose points of view differ. Of course, he checks the quality of his ·
> records: he gets program personnel to react to the accuracy of his
> portrayals; and audience members to react to the relevance of his
> findings. He does most of this informally - iterating and keeping a record
> of action and reaction. He chooses media accessible to his audiences to
> increase the likelihood and fidelity of communication. He might prepare
> a final report, he might not - depending on what he and his clients have
> agreed on. (p. 14)

Thus, the key question driving transactional evaluations is: What does the program look
like to the different people participating in it? Malcolm Parlett and David Hamilton
(1976), and Egon Guba and Yvonne Lincoln (1981) have written case studies using the
transactional model.

It is clear from this synthesis of formal models that evaluation is a complex
enterprise. Those responsible for conducting evaluations are often expected to do the
impossible with limited resources. Evaluators would need to have a wide repertoire of
knowledge and approaches, vast resources, teams of experts, and flexible timelines to
successfully meet the demands of most evaluative situations or to fully utilize some of
the existing evaluation methodologies. Triangulation, the combination of methodologies

in the study of the same phenomena or program, though highly desirable, is rarely practiced, perhaps due largely to the practical constraints just named and the high-level skills needed to conceptualize and carry out what Guba and Lincoln (1981) call "emergent designs." Yet, emergent designs are precisely the ones needed by the teacher-evaluator faced with the complex evaluative tasks that are a part of teaching.

Hence, evaluators of educational programs must make decisions within reasoned perimeters about which approach(es) to follow. Most opt for the one or several with which they are most comfortable and experienced. That is, they do what they do best. When a client contracts with an evaluator, the client buys not only the evaluator, but the evaluation style, methodological approach, academic training, and biases that the evaluator prefers. The experienced evaluator is careful at negotiating the evaluation contract so as not to be bound by services that she or he cannot or is unwilling to deliver. Even so, a host of political factors bear upon the evaluator, and methodological, design shifts can occur to accommodate what is politically feasible.

In sum, the meaning and practice of evaluation appears to have many dimensions. These dimensions are linked to the 1) purpose of the evaluation; 2) decisions to be made; 3) political concerns, values, issues, and audiences at stake; 4) expertise of the evaluator; and, 5) societal values and philosophical assumptions that influence the focus of evaluation and resulting decisions about a program's worth. It is wise for clients, therefore, to be thorough in thinking about five analogous questions in this regard: What is the purpose of the evaluation? What are the important short and long term decisions that will be forthcoming about the enterprise being evaluated? Who are the primary and secondary constituents of the program or enterprise and what are their perceptions about what questions and data permit meaningful evaluation? What is

the evaluator's "specialty?" Will this preferred approach satisfy our needs and the needs of our constituents?

Evaluation as a Function of Teaching

In this part, I wish to make the transition from understanding evaluation as an "outside-in" process of determining the worth of an educational program to an "inside-out" process of improving the quality of environments for learning. Carter V. Good, editor of the Dictionary of Education, 3rd edition, 1973 cites 36 definitions of constructs using the term evaluation other than program evaluation. The descriptions of approaches to program evaluation in contemporary practice reviewed in the second part of this section of the chapter provide a useful framework for teachers to create designs for evaluating their own instructional enterprises. Educational evaluation and curriculum evaluation will be briefly defined next, in order to deduce the evaluation of instruction, the construct aimed at in this discussion.

Educational evaluation. W. James Popham (1975) suggests that "systematic educational evaluation consists of formal assessment of the worth of educational phenomena." (p.3) These phenomena can include many things such as the outcomes of an instructional program, the instructional program itself, educational innovations, and the goals to which educational efforts are addressed. Popham's definition is helpful in getting us thinking about the purposes and potential of evaluation in education. From this definition alone, it can be inferred that there are other central uses in education to which evaluation could and should be put.

Curriculum evaluation. Several curriculum authors writing from what Pinar (1978) calls the "traditionalist" viewpoint have described the purposes of curriculum evaluation. Curriculum processes have their own built-in approaches to evaluation.

Hilda Taba's work (1962) is illustrative. Taba bases her definition of evaluation on certain assumptions about education. That is, education is a process which seeks to change student behavior. Education includes the mastery of information in connection with the study of certain subjects, but it also includes the reactions of students to this content. From these assumptions Taba frames her definition: "Evaluation is the process of determining what these changes are and of appraising them against the values represented in [curriculum] objectives to find out how far the objectives are being achieved." (pp. 312-313)

James Thornton and John Wright (1963) propose a similar definition of evaluation as a dimension of curriculum: "Evaluation is the estimation of growth and progress toward objectives or values of the curriculum." The primary purpose of evaluation is to give direction to future planning, to continue favorable practices, and to eliminate or correct less desirable aspects of the existing curriculum. Thornton and Wright explain that evaluation of curriculum must be based on evidence which, when interpreted, will show the degree to which objectives of the school program are being met (pp. 302-304).

It could be observed from these two examples that the model of evaluation one might subscribe to in program evaluation would probably influence other conceptualizations one holds of evaluation in educational practice. Taba and Thornton and Wright's definitions of curriculum evaluation are related to the assumptions, purpose and rationale underlying Ralph Tyler's behavioral objectives model.

J. Gaylen Saylor, William Alexander, and Arthur Lewis (1981) have conceptualized a model useful for showing the scope and nature of curriculum evaluation. Formative and summative (Scriven, 1967) data are emphasized and decision

parameters are suggested. The Saylor, Alexander, and Lewis model of curriculum evaluation helps to more clearly delimit what is referred to in this research as evaluation of instruction.

Evaluation of instruction. Despite all the varied and sophisticated experts, methodological approaches, and devices used for evaluation today, most of the responsibility for evaluating student learning resides with teachers as they aim at curriculum targets or objectives through instruction in classroom learning environments. Evaluating instruction means determining what pupils have accomplished and detecting what conditions have contributed to their success or failure.[13] Evaluation of instruction is concerned with the informal and more systematic techniques teachers use to assess student growth and achievement and to build successful learning environments. It is with this focus - the evaluation of pupil growth and accomplishment, individually and in groups - that this study centers its inquiry.

To conclude, these conceptualizations of evaluation advance the premise that the conscious and deliberate evaluation of instruction by teachers is a powerful entry point toward making classroom learning environments more responsive to all learners. The meanings of daily actions in school and classroom life as perceived by the actors themselves [teachers] have been virtually ignored in the standard approach to research on teaching. After all that has been written about it, we may know less today than we ever did about the meaning of evaluation in contemporary educational practice. Hence, an alternative explanation for the demise of educational evaluation as a promising field of inquiry may be that a significant "meaning gap" exists between scholars' writings and teachers' perceptions of evaluation. This gulf in understanding has had the effect of rendering a tide of prescription at district, state and national levels to a trickle of

inconsequence at the classroom level. If we wish to transform present cognitive structures of teachers toward the evaluation of instruction in an effort to realize its potential for contributing to increased equity and excellence in the classroom, then it is wise to first attempt to better understand how individual teachers view evaluation in relation to their own classroom practice.

Theoretical Perspective of the Study

It is clear from a consideration of alternative approaches to evaluation in contemporary educational practice that different evaluation theorists are guided in their work by different theoretical assumptions and accompanying methodologies. So too, have researchers on teaching been concerned with different theoretical assumptions and methods as they investigate the phenomena associated with classroom teaching. Thomas Kuhn (1962) has suggested that science undergoes a major revolution when it shifts from a dominant paradigm of inquiry to alternative paradigm(s) of inquiry. With upheavals in conventional ways of thinking fruitfully multiplying throughout the scientific community today, it is important for the researcher to situate his or her inquiry in an appropriate theoretical perspective(s). The discussion is organized into four parts. The first part defines an important construct in scientific thinking and sketches three frames of reference for viewing the world. The next three parts expand the three basic frames of reference into major research perspectives characteristic of present research on teaching. These research perspectives have been referred to by Bogdan and Biklen (1982); Erickson (1986); McCutcheon and Jung (1990), and others as 1) positivist; 2) interpretivist; and, 3) critical theory or critical science paradigms. The section concludes by suggesting that the present inquiry is guided by assumptions from the interpretivist and critical theory perspectives.

Scholars' Frame of Reference Revisited

A modest reflection on the history of human thought informs us that human beings have always been curious about the world they live in and have made rigorous attempts to understand it. The same basic questions of interest seem to prevail across eras and across cultures. What is the nature of the world? How can we come to know it? Who are we? What is the "truth" about these matters? Scholars assert that different philosophical orientations to thinking about these questions shape different responses to them (Meehan, 1969; Eichelberger, 1989).

As was illustrated in the previous section on the meaning of evaluation, many scholars have constructed terms and associated systems of ideas as a way of gaining leverage for understanding the complexity of the real world. Such a system of ideas that either "gives us some judgments about the nature of reality, or a reason why we must be content with knowing something less than the nature of reality, along with a method for taking hold of whatever can be known" (Reese, 1980 p. 352) has been referred to as a *paradigm* in the literature (e.g., Kuhn, 1962, Lincoln and Guba, 1985). In this discussion, the terms "paradigm" and "theory" are loosely constructed and represent similar notions.

Moreover, Lincoln and Guba (1985) suggest that inquiry in the physical and social sciences seems to have passed through a number of "paradigm eras," periods in which certain sets of beliefs guided inquiry in quite different ways. Scholars have described these paradigm eras in different ways as well. Following from Kuhn's (1962) notion of dominant paradigm periods, some scholars have referred to "pre-positivist," "positivist," and "post-positivist" periods of inquiry in categorizing western intellectual thought from Greco-Roman times to the present.[14] However, these labels seem to

reflect more the central and persistent dominance of positivist thinking than a substantive distinction among the periods.

Perhaps a more useful categorization of world view orientations is offered by Habermas in <u>Knowledge and Human Interests</u> (1972). Habermas claims that knowledge is produced by the ways people orient themselves to the world. He outlines three basic frames of reference, each of which is governed by a predominant interest. One is an orientation toward material well-being, governed by a technical interest in acting upon the world. This produces an empirical way of knowing manifest through facts and generalizations. A second orientation, toward communication, is governed by a practical interest in understanding others. This produces a situational and interpretive way of knowing manifest through the speech and meaning perspectives of the people being studied. The third orientation is toward freedom and it is governed by an emancipatory interest in freeing people from oppressive conditions. This produces a critically reflective knowledge. These orientations to the world, as described by Habermas, seem to correspond to the cross-disciplinary paradigms of inquiry termed positivism, interpretivism, and critical theory, respectively. Each of these paradigms and its major tenets are described next to provide the theoretical bases which underpin the current study.

Positivism

Reese (1980) has defined positivism as "a family of philosophies characterized by an extremely positive evaluation of science and scientific method." (p. 450) Positivists view the variables they are studying as actually existing in the universe. They are convinced (positive) that the relationships they "discover" have always been there. Positivism, in its most fundamental sense, assumes "not only that there is an external

world, but also that the external world itself determines absolutely the one and only correct view that can be taken of it, independent of the process or circumstances of viewing." (Kirk and Miller, 1986, p.14)

Positivism traces its roots through basic philosophical views of human nature held in western civilizations from ancient to more modern times. From Plato's idealism (i.e. - the nature of ideas is eternal and unchanging; this predetermined classification of ideas is accessible to the human mind; the physical senses are unreliable but may serve to reinforce what the mind already knows; in short, if we cannot conceive it, it does not exist) and Aristotle's realism (i.e. - there is one reality; the world operates on fixed natural laws; these laws are discoverable through observation and reason/deduction); to the empiricist ideas of Francis Bacon and John Locke who viewed the information that we garner from our sensory experiences in the world as most important for gaining knowledge. Later, positivistic assumptions were expressed in more complete systems of ideas that strike a balance between radical rationalism (the combination of idealism and realism) which distrusts human senses and uses only formal logic and reason to gain knowledge, and radical empiricism, in which knowledge is produced by inductive processes from human experience. Here the work of Immanuel Kant (The Critique of Pure Reason, 1787) and John Stuart Mill (A System of Logic, 1843) is exemplary.

There are significant assumptions underlying the positivist perspective that can be contrasted with those of interpretivist and critical science perspectives. Five such assumptions have been summarized by Lincoln and Guba (1985) as follows:

- An ontological assumption of a single, tangible reality "out there" that can be broken apart into pieces capable of being studies independently; the whole is simply the sum of the parts.

- An epistemological assumption about the possibility of separation of the observer from the observed - the knower from the known.

56

- An assumption of the temporal and contextual independence of observations, so that what is true at one time and place may, under appropriate circumstances (such as sampling) also be true at another time and place.

- An assumption of linear causality; there are no effects without causes and no causes without effects.

- An axiological assumption of value freedom, that is, that methodology guarantees that the results of an inquiry are essentially free from the influence of any value system (bias). (p. 28)

In addition, the positivist paradigm has assumed that the natural and social sciences have similar aims, that is, the discovery of general laws that serve for explanation and prediction. The methods for discovering general laws in the natural and social sciences are presumed to be similar as well. For example, positivists believe in preordinate research design and therefore hold that a preordinately designed research instrument is necessary. The same observable behavior should be sought for all subjects. Instruments are designed or selected that most appropriately test the hypotheses that frame the research questions. Data are collected under controlled circumstances from the number of subjects necessary for the research, positing that large samples suppress idiosyncracies (partial causes) and reveal general causes (the ultimate laws of nature).

Ironically, the result of mainstream positivist assumptions infiltrating research on teaching over the past 20 years in the form of process-product studies, may have succeeded in generating only partial theoretical models that are based primarily on the assumption that what can be found to be generic across classrooms would emerge across studies, and that the subtle variations across classrooms were trivial and could be accounted for in the analysis as error variance (Erickson, 1986). Hence, positivist perspectives have been criticized as inappropriate for generating a complete picture of

classroom teaching effectiveness. Instead, more interpretive studies with radically different underlying assumptions about the uniformity of nature and cause in social life that consider variations in classroom life more seriously, have been argued for as a more suitable way to build theory about teaching.

While the critiques against positivism have been severe, they have not rendered this traditional paradigm obsolete. Gage (1985) has argued that most controversy surrounding the issue of rival paradigms is spurious and has instead called for increased collaboration rather than disjuncture between, for example, ethnographic-sociolinguistic and process-product research paradigms on teaching.[15]

Interpretivism

In the last 25 years, a number of alternative approaches to research on teaching have been developed to provide a fuller understanding of the special character and quality of events that occur in classrooms. These approaches have been variously called qualitative, ethnographic, naturalistic, case study, participant observational, phenomenological, constructivist, symbolic interactionist, and interpretive. As with the evaluation approaches or models described earlier, these research approaches are relatively new in their application to education and there exists much disagreement among scholars employing these approaches over the proper conduct of their work and its theoretical foundations. Erickson (1986) uses the term *interpretive* to refer to the whole family of approaches that involves participant observational research. He adopts this term as an umbrella term because it comfortably accommodates interactive and phenomenological perspectives and hence, is a more inclusive term than the others. Further, it more strongly indicates the key feature that binds the various approaches

together - i.e. a "central research interest in human meaning in social life and its elucidation and exposition by the researcher." (p. 119)

Most of our present knowledge is based on *nomothetic* (abstract laws, generalizations) and analytical methods and procedures. Many educational researchers and social scientists question the value of the positivist's perspective when these theories and assumptions are applied to such human behavior as reading achievement, attitudes, interpersonal relationships, and the meaning that people attribute to their experiences (Eichelberger, 1989). In the interpretivist perspective, important knowledge about teaching is more *ideographic* (context-specific) in nature. Further, interpretivists believe that generalizations can be produced through careful compilation of "concrete universals," a paradoxical term borrowed from linguistics. That is, the systematic study of a specific case in great detail compared with other cases studied in equally great detail will gradually reveal universal properties of teaching which are present in each unique classroom context. These properties are manifest in the concrete, however, not in the abstract (Erickson, 1986). The shift in emphasis from a search for abstract, general laws to a search for concrete universals in research in teaching arises in part from the apparent failure of positivist perspectives to explain the tremendous variability and lack of stability of teacher effects on student learning. This anomaly emerges in spite of the goal of process-product studies, for example, to identify effective teaching practices that will reduce variability in student achievement. Carson (1990) has underscored the need for more interpretive inquiry:

> What is new in the debate today is the importance being placed on interpretation. In many fields of social science, including education, there is the sense that we have fragmented and abstracted human experience in such a way that we no longer understand it as it is embedded in life itself. In other words, we have been witnessing a deep and fundamental split between theory and practice. Thus in education, for example, we

have curriculum theories, theories of instruction, learning theories, and so forth each of which illuminate some aspects of teaching, but at the cost of removing these from their meanings in the lives of teachers and children. The turn toward interpretation is now an effort to reground our understandings in practice. (p. 172)

Interpretivism traces its foundations from phenomenology and hermeneutic philosophy. Important assumptions of this perspective are imbedded in these philosophies. Phenomenologists use human thinking, feeling, perceiving, and other mental and physiological acts to describe and understand human experiences. To understand the nature of human experience, phenomenologists believe that the experience itself must be studied and not an objective external world. A fundamental assumption is that human experiences can be catalogued and described in order to learn how meaning is derived from experience. Edmund Husserl, a founder of this philosophy (1858-1938), referred to those things that appear in our consciousness as "phenomena." Their study is called *phenomenology*. Grumet (1989) helps to clarify what phenomena are by suggesting that they include the things we think about that are "real" (chair, dog), "ideal" (peace on earth), or "fantastic" (unicorns and gremlins). Researchers in the phenomenological perspective attempt to understand the meaning of events and interactions to ordinary people in particular situations. According to Eichelberger (1989), phenomenologists do not assume they know what things mean to the people they are studying. They emphasize the subjective aspects of people's behavior. They attempt to gain entry into the conceptual world of their subjects in order to understand the meaning they construct around events in their daily lives. Phenomenologists believe that for human beings, multiple ways of interpreting experiences are available to each of us through interacting with others, and it is the meaning of our experiences that constitutes reality. Reality, therefore, is "socially constructed."

Hermeneutics is a greek term that refers to the art and science of interpretation (Reese, 1980). It is an ancient technique used for interpreting meaning from legends and other texts. This method has been used by some scholars to interpret (and re-interpret) the Christian Bible, for example. *Hermeneutic philosophy*, a more modern adaptation developed by Wilhelm Dilthey and other German philosophers, is the study of interpretive understanding, or meaning. Scholars from various social science disciplines have adapted the tenets of phenomenology and hermeneutics to develop various sub-theories in interpretivism. For example, scientific inquiry in the fields of social and anthropology has influenced the development of *symbolic interactionism*, a powerful theory in the interpretivist orientation that attempts to explain human behavior in terms of meanings (Bogdan and Biklen, 1982).

Much of the turn toward interpretivism stems from an historic awareness that traditional research approaches often excluded from consideration those people and groups in society who had little political power and hence were not often heard. Bogdan and Biklen (1982) suggest that, in sociology, the roots of symbolic interactionism can be traced from Frederick LePlay's (1879) participant observation studies of urban working class families in France, Henry Mayhew's life histories of working and unemployed poor people in London, (1851-1862),[16] and later in the United States, Miriam Komarowsky's (1946) study of women in higher education. Also, John Dewey's writings in pragmatist philosophy (e.g. 1910)[17] and interdisciplinary dialogue with his colleagues Charles Horton Cooley, Robert Park, and George Herbert Mead of the renowned Department of Sociology at the University of Chicago in the 1920s were influential in advancing the theory of symbolic interactionism.

In anthropology, many scholars operate from a phenomenological perspective in their studies of education. The theoretical framework for anthropological studies centers around the researcher's concept of culture. For example, Spradley (1980) defines culture as "the acquired knowledge people use to interpret experience and generate behavior." (p.6) This definition is congruent with symbolic interactionism's emphasis on meaning. Spradley suggests that the cultural knowledge that people have learned as members of a group cannot be observed directly. Cultural knowledge is explicit and tacit. Three types of information can be used to make inferences about culture. One can observe what people do (cultural behavior); observe things people make or use (cultural artifacts); and or listen to what people say (speech messages). Both explicit and tacit cultural knowledge are revealed through speech, whether in casual comments or lengthy interview. An important assumption in symbolic interactionism and in Spradley's work is that language is the primary means for transmitting culture, hence much of what can be known about any culture is encoded in linguistic forms (Spradley, 1979; 1980).

If meaning is central to theories of culture then it becomes necessary to study meaning carefully. The attempt to describe culture or aspects of it is called *ethnography*. In the interpretivist paradigm, the goal of ethnography is a search for cultural meaning. Bronislaw Malinowski's extended fieldwork in New Guinea and subsequent writings (e.g. 1960), and Margaret Mead's attempt to suggest implications of her work in Samoa for schooling in the U.S. (1951), were noted attempts to explicate theory in terms of their interpreted meaning of culture.

More recently, scholars have begun to carve further into cross-disciplinary theory building in research on classroom teaching. A group of scholars branching from the

interpretive tradition and the premises of symbolic interactionism view classrooms largely as communicative environments. Drawing on the fields of cognitive anthropology and sociolinguistics, some of these scholars have been referred to as *constructivists*, with their emphasis on context and meaning in classrooms as socially constructed events (e.g. Green, 1983; Puro and Bloome, 1987). An important notion for constructivists is that life in classrooms is holistic, continuous and intertextual. To understand the holistic nature of classroom life, constructivists believe that it is important to explore the interrelated nature of classroom events and the continuity of experience for learning in classrooms (Edwards and Mercer, 1987). One way to view continuity of experience is as "intertextuality." Intertextuality has been described by David Bloome (1989):

> Whenever people engage in a language event, whether it is a conversation, the reading of a book, diary writing, etc., they are engaged in intertextuality. Various conversational and written texts are being juxtaposed. Intertextuality can occur at many levels and in many ways.
>
> Juxtaposing texts, at whatever level is not in itself sufficient for intertextuality. Intertextuality is a social construction. The juxtaposition must be interactionally recognized, acknowledged, and have social significance. In classrooms, teachers and students are continuously constructing intertextual relationships. The set of intertextual relationships they construct can be viewed as constituting a cultural ideology, a system for assigning meaning and significance to what is said and done for socially defining participants.

Thus, the various theories evolving under the umbrella of interpretivism can be seen to converge around the following key assumptions, particularly as they apply to everyday life in schools:

- Human beings create meaningful interpretations of the physical and behavioral objects that surround them in the environment. Thus, meaning is constructed.

- School is a social setting where people construct and conduct life together.

- Formal and informal social systems are intertwined, thus to be able to interpret what occurs within any given classroom requires an understanding of that classroom as a mini-society with cultural norms and expectations, rights and obligations, and roles and relationships for its members.

- These microcultures differ from classroom to classroom. Careful and systematic elucidation of the particular leads to more general statements about the nature of teaching.

- The relationship between participant and researcher is natural, unobtrusive, conversational, and non-threatening; research methods are qualitative, flexible, designed to gain accurate reflections of participant views and interpretations of meaning.[18]

In summary, ethnographic-sociolinguistic research views teaching as an instance of symbolic interaction, the process whereby people act on the basis of the meanings events have for them. These meanings derive from social interaction. People develop and change these meanings in terms of the situation and their own actions (Bolster, 1983). Significant knowledge about life in classrooms or any other social context consists of the meanings that participants develop about that context or situation. The information gained from systematic inquiry in the interpretive mode can be used by teachers to illuminate and strengthen their own pedagogical practices. A growing number of educators are applying interpretive theoretical assumptions, grounded in other disciplines, to the systematic study of the teaching process.

Critical Science

Although the principles of equity and equal opportunity are important ones in a democratic society, many groups of citizens may not experience the full benefits that accrue for the majority. Poor people, people of color, women, people with unorthodox religious beliefs, and people with mental or physical handicap may face persistent discrimination. There have always been people who have resisted dehumanization,

spoken out against an unjust social order, taken action for equal rights. The teachings and writings of Jesus Christ, Frederick Douglas, Mahatma Ghandi, and Martin Luther King, for example, reflect critical attitudes toward the dominant socio-political system during their respective lifetimes. At the same time, their discourse implored all human beings to strive toward a higher, nobler more peaceful order.

The intellectual positions adopted by mainstream scholars in a society often become imbedded in the values and norms that perpetuate the status quo. Critical theorists argue that the status quo is never satisfactory. There is always room for more improvement. Like the interpretivists, critical theorists believe that knowledge is socially constructed and that there are multiple interpretations of reality. The central premise for these scholars, however, is that there are dominant constructions of reality that promote inequities. Institutions in society serve to perpetuate an imbalance of wealth and power among those with money and influence. The rich get richer, the poor get poorer. In our present society, one could argue that the politics of discrimination for many groups of people are alive and well in virtually every aspect of life - business and industry, health care, education, housing, legal system, government, and leisure and recreation. Needless to say, individuals who are insistent in reconstructing ways of thinking, acting, and enfranchising people toward greater equity and emancipation through these societal institutions face continuing political struggle.

Scholars from a variety of disciplines in the social sciences who adopt the perspective of critical science view their work as going beyond explaining, describing, or interpreting the meaning of events to calling for fundamental changes in society. Different scholars have marked off different boundaries in which to focus their critical analysis. Neo-marxists, for instance, have developed extensive critiques of capitalism and

the social and economic inequities that it perpetuates (e.g. Bowles and Gintis, 1976). Existentialist philosophers have advanced alternative conceptualizations to unseat the behaviorism that has dominated learning theory in education for decades (e.g. Pinar, 1975). The scholarship on women that has steadily grown in the 1980s has given rise to feminist theory (e.g. Grument, 1988). Meanwhile, some of the most penetrating critical analyses, are represented in critiques of American education that have suggested that the processes of schooling have systematically served to widen gaps in literacy among children (e.g. Illich, 1972; Katz, 1975; Apple, 1982; Meier, et al., 1989) rather than functioning as the ultimate equalizer as traditional advocates have proclaimed.

Praxis is the central concept for the critical theorist. *Praxis* derives from the greek term meaning "to do." By praxis, critical theorists mean what McCutcheon and Jung (1990) have referred to as "the emancipatory interplay between action and reflection." The authors go on to explain:

> Action by itself is directionless and reflection by itself aimless. The dialectical movement between action and reflection takes into account the complexities of the practical, sociocultural factors and the construction of meaning. This dialectical movement then makes possible the uncovering of basic inequities, which in turn makes possible a movement to emancipation. (p. 147)

Thus, the author writing from a critical perspective would not only study a problem of interest, but after careful analyses, would seek or recommend a series of reforms to change existing conditions that are perceived to perpetuate the problem. One way to more clearly understand this research perspective is to illustrate an example. The recent study by Meir, Stewart, and England (1989) is exemplary of critical inquiry around the issue of equal access to educational opportunity for black youth in American schools. The study shapes a political theory of second generation discrimination against

blacks in elementary and secondary education. The research examines equal educational opportunities in 174 U.S. school districts with at least 15,000 students and 1% black enrollment. It is premised on the view that education is the single most important civil rights issue affecting blacks in the United States. From this point of view, the authors suggest, therefore, that equal educational opportunity is the key to good jobs, quality housing and political influence.

Findings from the study indicate that patterns of inequity and discrimination accrue to children in school on the basis of race and class. For example, black youth in schools are found to be disproportionately sorted into lower academic groups. Black youth in schools more frequently experience corporal punishment, suspension, and explusion. Black youth drop out of school in disproportionate numbers to white students.

Meier, et al. then recommend substantive policy changes that they believe would ameliorate these and other hindering educational conditions for black children. Seven policy recommendations are made:

1. Abolish at-large school district elections to provide greater opportunities for minority representation.

2. Increase significantly the number of black teachers.

3. Transform the federal Office of Civil Rights from a mediocre data collection agency to an aggressive civil rights action agency.

4. Empower the federal Equal Employment Opportunity Commission to release aggregate data on racial employment by individual school district.

5. Eliminate or severely curtail the use of ability grouping.

6. Ban corporal punishment.

7. Review and restructure school disciplinary procedures.[19]

It can be implied from the study by Meier and his colleagues that inquiry from the critical perspective involves a conscious effort to re-examine institutionalized constraints of schooling that discourage rational and reflective dialogue between educators and students. Events are understood in terms of political, social, and economic hindrances to equal opportunity. Methodology is flexible but is aimed at raising questions which increase awareness of societal injustices and which suggest directions for constructive change. Table 1 summarizes some of the contrasting beliefs and assumptions of the positivist, interpretivist, and critical science research perspectives.[20]

In summary, scholars' conceptions of the world and the rules of evidence for "legitimate" interpretation of information have changed considerably over time and will no doubt continue to change. The purpose of this section of the chapter has been to surface some of the important epistemological assumptions being made in the present inquiry about the nature of reality, how human beings derive knowledge, and the methods they use to gather, analyze, and interpret information. It has been argued that the positivist paradigm alone has been inadequate for investigating the complex phenomena associated with classroom teaching and learning.

The present study is anchored in a blend of assumptions culled from interpretivist and critical science perspectives. A key premise maintained throughout the present research is that gaining knowledge about what teachers know about classroom grouping and evaluation is important. Inferences about what teachers know about these symbolic aspects of equity and excellence in classroom life are made from what teachers say. Some inferences are also made on the basis of what teachers use (i.e. instructional artifacts such as student work samples and teacher record-keeping documents). The

Table 1

Contrasting Beliefs and Assumptions of Three Different Research Perspectives

BELIEFS AND ASSUMPTIONS ABOUT:	POSITIVISM	INTERPRETIVISM	CRITICAL THEORY
ONTOLOGY			
The nature of reality	Single, can be quantified, measured, and categorized	Multiple, holistic, constructed through human interaction	Constructed, multiple, but there are dominant constructions that promote inequities
The relationship between the knower and the known	Separate	Interrelated, dialogic	Interrelated, embedded in society
EPISTEMOLOGY			
The nature of knowledge	Events are explained in terms of real causes or simultaneous effects; the right methods produce legitimate knowledge	Events are understood through active mental work, interactions with external context, transactions between one's mental work and the external context	Events are understood in terms of social and economic hindrances to true equity
The role of value in research	Value-free	Value-bound	Related to values of equity
METHODOLOGY			
The purpose(s) of research	PREDICT - Discover laws underlying reality	UNDERSTAND - what occurs and the meanings people make of phenomena	EMANCIPATE - Uncover and understand what constrains equity and supports hegemony to free oneself of false consciousness and change practice toward more equity
Procedures of inquiry	Observation; a priori procedures, theories; test and confirm hypotheses	Observation; interviews; participants help structure the inquiry; emergent design	Observation; interviews; methodology raises questions to heighten awareness of injustices and begin the change process

69

research seeks to change approaches to academic grouping and evaluation that may be hindering the full development of children.

The philosophical foundation of knowledge developed in this section of the chapter is applied next in the analysis of selected studies culled from literatures that have important implications for the major research questions that guided the present study. This review of literature is intended to be exemplary rather than comprehensive for each of the following veins of inquiry.

Selected Studies of Ability Grouping and Reading Instruction

Grouping children by ability is so pervasive that many teachers do not question its validity as a necessary condition for learning. Substantial research has been produced over the past 75 years to determine the effects of ability grouping in general on different populations of students. The overall results appear to be inconclusive. (For differing and often conflicting results see, for example: Otto, 1930; Cornell, 1936; Borg, 1965; Heathers, 1969; Rist, 1970; Stallings, 1978; Kulik and Kulik, 1982; Brophy and Good, 1987).

What is clear from the reported studies, however, is that while tracking and ability groups have been rationalized as means for adapting curriculum and instruction for diverse groups of students, there is no persuasive evidence available to suggest that these grouping arrangements improve achievement and are equally beneficial for low and high achieving students. It has been long assumed that grouping children by ability:

- promotes achievement - (students could advance at their own rate because they were grouped with other students of similar ability).

- provides students with more appropriate materials and content - (teachers could plan more challenging instruction for high-ability students while providing low-ability students more opportunities for success).

- spares less capable students the harm of embarassement and other damaging effects by not forcing them to compete with brighter students - (rather, students are challenged to do their best in a more realistic range of competition).

- Makes teaching easier - (because it narrows the tremendous range of student diversity existing in any classroom).[21]

It appears now that many of these historic assumptions about ability grouping are unraveling. In fact, more recent research has implied contrary conclusions (e.g. Oakes, 1985; Slavin, 1987; Meier, et al., 1989 and the studies further reviewed in this section). The following generalizations about ability grouping are implied from cumulative studies of the past decade:

- Ability grouping has few short-term effects on achievement as measured by standarized tests.

- Grouping students by ability seems to have a negative effect on the self-concept and self-esteem of lower-tracked students.

- By separating some students into instructional groups of differing achievement levels, grouping can have the indirect effect of isolating or resegregating students who differ in social class, race, or ethnicity.

- Especially for lower achievers (due in part to self-fulfilling prophecy effects, less engaged learning time, and mismatches in instruction with how students learn) ability grouping may actually lead to increased differences in academic ability and achievement among students.

- Ability grouping can predetermine students' future opportunities and aspirations.

- Instruction in ability-grouped classes may be more difficult than in heterogeneously classrooms since ability grouping has little impact upon the real heterogeneity of a class, and since management problems seem to exacerbate during instruction for lower tracked students.

- Ability grouping is an ineffective means of addressing individual differences.[22]

These are significant findings that have profound implications for the egalitarian principle of equal educational opportunity.

Seven studies were reviewed and analyzed in terms of their essential purpose, focus, theoretical framework, methodology, and findings. The studies were selected on the basis of their contemporary time frame (last 10 years), representativeness of research on ability grouping and reading instruction over this time period, and diversity in research perspective taken to investigate the problem. Table 2 helps to provide a meaningful comparison of the studies by Allington (1980, 1982); Borko and Eisenhart (1986); Eder (1981); Grant and Rothenberg (1986); Hallianan and Sorenson (1985); and Rown and Miracle (1983), respectively.

Results of these studies show that teachers may treat children in different groups differently, and that on many dimensions of teacher behavior, students in higher level groups appear to be treated more favorably during instruction than are students in lower groups. Moreover, the structure of ability grouping has been described as contributing to the formation of distinct literacy communities within schools and classrooms which foster substantial differences in the way text is used and in the way attitudes and expectations toward reading are shaped. For these reasons, it has often been concluded that the perpetuation of ability grouping widens differentials in achievement between higher and lower groups, and constrains peer relationships to interactions within but not often among the groups. Further, the accuracy of assessment procedures used to group students by ability has been questioned since the results have sometimes created cleavages of students in classrooms on the basis of race and social class. Once assigned to initial reading groups, student mobility across groups tends to be infrequent.

The results of this selected group of studies leads to an overall conclusion that the pervasive use of ability grouping, even as a means for organizing students for reading instruction, may be at odds with what we have learned about motivating and

Table 2

Comparison of Research Elements: Selected Studies of Ability Grouping and Reading Instruction

Study	Purpose	Foci	Theoretical Orientation	Methods	Findings
1. Allington (1980)	To determine whether there are differences in instructional environments of good and poor readers	Teacher interruption behaviors during primary grade oral reading	Positivistic (Process-Product)	Observations, 21 first grade classrooms	• There are substantive differences in the reading instruction provided for good and poor readers, particularly in the amount (words read per minute) and mode (silent vs. oral) of reading assigned during group work.
2. Allington (1982)	To determine whether there are differences in instructional environments of good and poor readers	Quantity and mode of assigned reading during group instruction.	Positivistic (Process-Product)	Audiotaped reading lessons, questionnaire for 20 first and second grade teachers	• Teachers are more likely to interrupt miscues of poor readers, regardless of the semantic appropriateness of the error.
3. Borko and Eisenhart (1989)	To describe differences in ability-based reading groups as differences in classroom literacy communities	Uses of print; ways of communicating about print; students' behaviors and understandings related to literacy; teachers' expectations about student performance; standards of literacy competence.	Interpretive (Ethnographic)	Observation, interviews, four second grade teachers and four focal students, one from each of four ability-based reading programs in each classroom	• The content of reading instruction, student conceptions of reading, rules for participation in literacy events and teacher expectations and evaluations of students differ across ability groups, suggesting that ability-based reading groups can be conceived as distinct literacy communities.

Table 2 cont.

Study	Purpose	Foci	Theoretical Orientation	Methods	Findings
4. Eder (1981)	To examine the nature and extent of differences in teacher-student interactions in ability-based reading groups.	Differences inattentive student behavior, teacher management, and reading turn disruptions between high and low ability groups.	Process-Product (Sociolinguistic)	Observation; one first grade classroom over an entire academic year; videotaped reading group lessons of four within-class ability groups.	• Learning contexts of lower ability groups are likely to differ from learning contexts of higher ability groups in crucial interactions between teacher and student during reading instruction. Lower groups have greater inattention, more management activities and more reading turn disruptions than higher groups which are likely to contribute to lower levels of reading achievement for students assigned to low ability groups.
5. Grant and Rothenberg (1986)	To compare student experiences in different ability-based reading groups.	Effects of within-class ability grouping on opportunities of students to acquire academic skills and develop social relationships with teachers.	Critical (Sociology)	Secondary analysis of ethnographic data; 8 first and second grade classrooms	• Students in top groups are more advantaged than students in lower groups in the following ways: - They work in environments that are conducive to learning more academic skills - They have more opportunities to demonstrate competence

Table 2 cont.

Study	Purpose	Foci	Theoretical Orientation	Methods	Findings
					- They practice autonomous, self-directed modes of learning - They have different social relationships with teachers
6. Hallinan and Sorensen (1985)	To determine the effects of ability grouping on children's social relationships	HO: membership in the same ability group has a positive effect on the choice of a peer as a close friend for elementary school children.	Positivistic (Hypothetico-analytic)	Correlation: Density analysis, clique analysis, and logit analysis 104 reading ability groups in 32 classrooms, 10 elementary schools.	• All three methods of analysis support the conclusion that membership in the same ability group increases the likelihood that students will become best friends. Ability groups become more cohesive and more attractive to their members over time, but may foster a stratified friendship network that may further encourage race and social class cleavages in the classroom.
7. Rowan and Miracle (1983)	To analyze the effects of ability grouping on achievement.	Effects of differential peer influence and differential instruction on reading achievement in within-class and between-class structure of ability grouping	Positivistic (Process-Product)	Systematic classroom observation, analysis of school records; 6 elementary schools, 10 fourth grade classrooms, 200 students	• Group rank affects the way students are taught, however, there is no clear evidence that the pattern of differential instruction reinforces initial inequalities in achievement

75

Table 2 cont.

Study	Purpose	Foci	Theoretical Orientation	Methods	Findings
					• Group rank stratifies peer ability contexts and peer ability contexts affect student behavior, however, there is no clear evidence that achievement differentials in ability groups are due to differential patterns of peer influence

increasing the reading achievement of children (Carbo, 1987). Rather, than helping to minimize initial disadvantages in learning, ability grouping increases these disadvantages. Given the serious implications for the academic, social, and affective development of learners that research on ability grouping poses, and given the crucial importance of a solid foundation in reading in the elementary curriculum it is crucial to discover the criteria teachers use to place students into groups for classroom reading instruction. It follows that these criteria should also be carefully considered by teachers. Are they? The present inquiry attempted to provide some useful data and insight for addressing this question.

Selected Studies of Classroom Pupil Evaluation

A general review of literature on research in classroom pupil evaluation reveals that the dominant paradigm has been limited, focusing on testing as the primary means for determining student achievement. This research perspective, inherited from the strong positivistic influence of educational psychology and measurement, presents only a small portion of the range and quality of evaluation that takes place in classrooms. Because of the narrow scope of measurement research, little is known about the nature, role, and quality of evaluation developed and used by the classroom teacher (Stiggins, Conklin, and Bridgeford, 1986).

This conclusion has been confirmed by other reviewers of research on testing in the schools. For example, Shulman (1980) concludes that "In general, the kinds of tests we use are inconsistent with, and in many cases irrelevant to, the realities of teaching" (p.69). Similarly, Lazar-Morris et al. (1980) suggest that:

> In-class assessments made by individual teachers have yet to be examined
> in depth. How these and other assessments are united with teacher
> instructional decision-making processes and how they affect classroom
> organization and time allocation to other objectives are areas that should

be explored. Teachers place greater reliance on, and have more confidence in, the results of their own judgments of student performance, but little is known about [these] kinds of activities. (p. 24-25)

Although research on classroom evaluation has tended to focus on the role of standardized tests, a few scattered studies have provided modest insight into the nature of classroom evaluation. The following are general findings, reported by Stiggins et al. (1986) and summarized previously by Airasian (1984):

- There are essentially two sets of characteristics evaluated in the classroom: academic and social. (Kellagan and Airasian, 1982)

- The relative importance assigned to these two factors varies with grade level, with social factors seen as more important in elementary school. (Salmon-Cox, 1981)

- Teachers "size up" students as individuals, group them very quickly, often considering background and family variables over achievement indicators. These initial estimates remain quite stable. (Rist, 1970; Salmon-Cox, 1981)

- Students appear sensitive to these early teacher assessments, learn their position in the "pecking order" of the class and respond accordingly. (Rist, 1970)

- Teachers interact differently with students they perceive to be of high or low ability. (Brophy and Good, 1986)

- Regular classroom teachers and reading specialists can accurately predict student test performance, but have trouble diagnosing student reading difficulties accurately. (Weinshank, 1980)

Further studies on the relative value teachers attach to different types of assessments indicate that the ones considered most important for classroom purposes originate with the teacher (Stiggins and Bridgeford, 1985). Available evidence also suggests that observation and subjective judgment are important evaluative tools used by the teacher-evaluator (Salmon-Cox, 1981). Many teachers also apparently rely on their

own assessments as the primary source of information on student progress and achievement (Herman and Dorr-Bremme, 1982).

Since there is such scant evidence on the nature of classroom evaluation as conducted by teachers, not much can be reported about the quality of the evaluation methods used by teachers. Several studies from the research on testing in the schools literature, however, have found that there is a need for a broader range of testing approaches and improved test questions on teacher-developed assessments (eg. Fleming and Chambers, 1983).

In short, a general review of the literature on classroom pupil evaluation shows that the classroom evaluation tasks of the teacher range from diagnosis to grouping to determining learning progress to grading and reporting results to parents. These functions of classroom evaluation are extremely complex and represent different types of assessment demands and different types of decisions by the teacher. To date, few studies detail the broader range of evaluation activities of teachers that include observation, oral questioning, performance assessment, grading, marking, and the like.

Three studies on classroom pupil evaluation have been selected for further analysis and comparison with the present study. These studies were selected for their contemporary relevance, their thrust in new and recommended directions in this branch of inquiry, and their points of merger with the present study. Table 3 provides a useful basis for comparison of the purpose, focus, theoretical orientation, methodology, and findings of the Barnes (1985), Stiggins and Bridgeford (1985), and Thiessen and Moorhead (1985) studies, respectively.

Results from these selected studies show that the research agenda for evaluation of student learning has broadened to include a more comprehensive description of

Table 3

Comparison of Research Elements: Selected Studies of Classroom Pupil Evaluation

Study	Purpose	Foci	Theoretical Orientation	Methods	Findings
1. Barnes (1985)	To describe teacher behaviors along selected dimensions of classroom pupil evaluation	Teacher attitudes, knowledge, and processes used to carry out pupil performance evaluation in the classroom	Interpretive (inferred) (qualitative approach)	Descriptive via journal entries, audio-taped conferences, interviews, and observation; 20 cooperating teachers (CTs) and 20 student teachers (STs)	• STs use similar informal, continuous evaluation processes as their CTs • CTs use primarily intuitive means to determine "where students are" • Parent conferences are seen as an important element of the evaluation process by CTs, less so by STs • Clear criteria for the evaluation of pupil learning process not articulated by CTs nor STS • Pervasive and unresolved conflict among CTs and STs concerning evaluation on the basis of performance and evaluation on the basis of effort.
2. Stiggins and Bridgeford (1985)	To broaden understanding of the classroom assessment environment	Nature and use of teacher-developed assessments (i.e.,	Positivistic (inferred) (preordinate design)	Structured questionnaire; national survey of stratified sample of	• As grade level increases teachers report using their own tests more than

Table 3 cont.

Study	Purpose	Theoretical Orientation	Methods	Foci	Findings
			teachers from eight school districts differing in size and geographic location; 384 teachers surveyed, 228 questionnaires completed and returned	Teachers own paper and pencil tests and methods of performance assessment); also, teachers' concerns about assessment	published tests to assess student learning. • Teachers report being concerned and wanting to improve their assessment methods but do not appear to be in the process of changing the ways that would improve them • Performance assessments of teachers have technical flaws that may invalidate their results
3. Thiessen and Moorhead (1985)	To discover the meaning and reasons for the judgments and practices teachers use in student evaluation	Interpretive	General survey of 450 teachers followed by more intensive interviews with 30 teachers	Teachers responses to two major questions: how should evaluation influence the learning of each student? What guidelines should govern the evaluation of students?	• Teachers express student evaluation concerns as issues; teachers differ according to which practices they adopt to resolve the issues. • Some teachers prefer practices which emphasize student self-enhancement, diagnose both their teaching and student learning, focus on the total student, build accountability from within

81

Table 3 cont.

Study	Purpose	Foci	Theoretical Orientation	Methods	Findings
					their classroom and vary the formality of the evaluation approaches they adopt each day. (Interactive orientation)
					• Other teachers prefer practices which emphasize student self-development, focus on what and how their students think, maintain accountability to their school board, the parents and the community and vary the type of formal evaluation approaches they regularly schedule for their students. (Responsive orientation)
					• Similar evaluation practices have different meaning and importance for teachers. The teacher's orientation or philosophy toward evaluation is what distinguished one teacher from or with another rather than specific student evaluation practices.

classroom assessment environments. The Stiggins and Bridgeford study (1985) is helpful in providing data that describe in greater detail the importance teachers attach to different forms of assessment as purpose, grade level, and subject matter vary. The study also surfaces some of the expressed concerns of teachers around the use of their own assessments. Unfortunately, Stiggins and Bridgeford's definition of assessment seems to be restricted to teacher-developed tests. As a result, interpretation of findings is limited to the importance of different types of *tests* and the concerns teachers have regarding the quality of the *tests* they use to assess student learning. This study invokes a highly structured questionnaire as the primary research instrument and quantitative data analysis. Consequently, the lofty purpose of the study to broaden understanding of the classroom *assessment* environment is only partially realized.

Barnes' study (1985) is an attempt to link teachers' attitudes, knowledge bases, and practices in classroom pupil evaluation to perceived neglects in teacher education curriculum. An important contribution of this study is its multi-method approach to data collection and description of common themes among teachers and interns regarding student evaluation. Interview data reported for this study suggest that teachers intuitive judgments of "where students are" may play a predominant role in classroom pupil evaluation. However, neither cooperating teachers nor student teachers articulated clear criteria for evaluating student learning progress. Moreover, both groups of teachers expressed a continual internal conflict in basing student evaluation on achievement versus effort.

Interestingly, Barnes reports that the two groups of teachers had different conceptions of the purpose of pupil evaluation. For most cooperating teachers, evaluation had one main purpose, to provide a source of information for parents and

students in the form of grades. Student teachers on the other hand viewed evaluation in terms of four major functions: a) as a strategy for motivating students; b) as a communication link to parents; c) as a way to classify and group students; and, d) as a method to evaluate their own effectiveness. These more diverse and substantive conceptions toward evaluation may be attributable, in part, to the recent formal preparation received by this group of teachers, contradicting one of the major themes of this study. As with the Stiggins and Bridgeford study, the described purpose of the Barnes study falls short of its full intent. Reported findings provide no insight into the specific criteria teachers use to evaluate student learning and little insight into the broad array of evaluation methods teachers use to determine student's learning progress.

The Thiessen and Moorhead study (1985) is unique in the theoretical approach used to investigate teacher practices in student evaluation. The study adopts an interpretive perspective which seeks to understand the meaning of classroom evaluation from listening to what teachers have to say. One question was asked of all participants: "When you think about student evaluation, what are you concerned about?" From an analysis of teacher responses, Thiessen and Moorhead identified 17 concerns across five specific issues. To resolve the student evaluation issues, the researchers say teachers adopt practices which are sensitive to the particular conditions of their classroom situation and to the uniqueness of their students. The teachers differed in the practices they preferred to resolve the issues. Thiessen and Moorhead's conceptualizations of teachers with *interactive* student evaluation orientations and *responsive* student evaluation orientations represent significant theoretical propositions worthy of further study.

In sum, several studies in classroom pupil evaluation in the mid 1980's contributed to a more comprehensive understanding of classroom evaluation

environments than most of the studies before them. Researchers of these descriptive studies have called for more qualitative studies on classroom evaluation practices. One missing element is a richer description of the specific methods and performance criteria teachers use to determine student learning progress in a particular subject matter area. In addition, the research on ability grouping and classroom evaluation can be merged around an investigation of whether teacher evaluation methods and performance criteria differ for students in different instructional groups within the classroom. The present study attempted to provide useful information for answering this question.

Selected Studies of Teacher Decision-Making

An important premise of the present study is that too little attention has been given to understanding complex issues in student evaluation that teachers face daily. Available research on ability grouping and classroom pupil evaluation provides some insight about the nature of classroom evaluation environments, however, the overall research picture still lacks focus and detail. To gain a more elaborate view of classroom evaluation, one can find clues imbedded in other distinct areas of research on teaching.

A timely vein of research on teaching has been developed during the last 15 years that attempts to examine not only teachers' behaviors but also their judgments, plans, and decisions. This research provides a theoretical and empirical base for examining Research Question 3 of the present study which seeks to illuminate how teachers use the information emerging from their classroom evaluation of pupil learning progress in Reading. Research on teacher planning and decision-making rests on two important assumptions that are also adopted in the present study: 1) teachers behave reasonably in making judgments and decisions; 2) teachers' behavior is guided by their thoughts, judgments, and decisions (Shavelson, 1983).

Clark and Peterson (1986) have organized the research on teacher decision-making into three main categories under the general heading of teachers' thought processes. The three categories are: 1) teacher planning, 2) teacher's interactive thoughts and decisions, and 3) teachers' theories and beliefs. Teacher judgment is viewed as a cognitive process that runs across all three categories. Taken together, these studies have begun to build a more complete view of teaching and its many interrelated functions. The research on teacher decision-making has often focused on three corresponding types of decisions, each placing significantly different evaluative demands on teachers. These are: preinstructional decisions (preactive planning), interactive decisions, (made during instruction), and postinstructional decisions.

Four studies, representing the different decision contexts, were reviewed for more intensive analysis. Table 4 facilitates comparison of the purpose, foci, theoretical orientation, methodology, and findings of the studies by Doron Gil, Paula Stern and Richard Shavelson, Sylvia Pratt Whitmer, and Robert Yinger, respectively.

The Yinger (1980) study focuses on preactive planning decisions of teachers. Studies of teachers' preactive decisions suggest that teachers are interested in getting to know their students as individuals. At the beginning of the year, they collect a considerable amount of information about their students' academic and social characteristics. Available evidence indicates, however, that teachers tend not to focus on these student characteristics when planning for the instruction of the class. Instead, planning decisions are built primarily around the goal of gaining and maintaining student cooperation in classroom activities (Yinger, 1977). A significant implication of Yinger's work is that the implicit theories of teachers - i.e., their beliefs about cause and effect relationships - are largely theories about cooperation rather than learning.

Table 4

Comparison of Research Elements: Selected Studies of Teacher Decision-Making

Study	Purpose	Foci	Theoretical Orientation	Methods	Findings
1. Gil (1980)	To describe teachers' classroom diagnostic practices	Instructional decision-making, diagnostic and remedial practices in reading.	Positivistic (inferred) (Clinical/prescriptive, one-best-way approach)	Descriptive, multi-method; laboratory simulations, product measures, process measures, classroom observations, and interviews for 10 teachers	• Teachers showed little agreement with each other on their stated diagnostic judgments of children's reading difficulties • Teachers differed markedly in the processes they used to diagnose a case - i.e., length of interaction, number of cues collected, number of final stated diagnoses and number of comments made • Teachers mention similar but vague diagnostic categories from which to base their decisions about student reading problems
2. Stern and Shavelson (1981)	To explore the relationship among teachers' judgments, decisions, and instructional behavior in reading	How judgments of students' abilities influence their grouping decisions for reading instruction, and how in turn, grouping decisions determine certain instructional behaviors	Interpretive (Ethnographic approach)	Analysis of written observations of a combined grades (5th and 6th) classroom during small group reading instruction for a six month period; also structured and unstructured interviews;	• Teachers group students for reading on the basis of ability, even though they have access to many sources of information about their students.

Table 4 cont.

Study	Purpose	Foci	Theoretical Orientation	Methods	Findings
				teacher descriptions of each student in the class; and, student and teacher work samples. (N=2 teachers, twenty students)	• Once grouped the group and not the individual student became the primary unit for instructional planning and decision-making. • Teachers' plans for the low ability group differed considerably from their plans for the high group.
3. Whitmer (1983)	To develop an understanding of teacher marking processes	Teacher thoughts, judgments, and decisions during the marking task	Attribution and utility theory	Structured interviews with five experienced upper elementary teachers; qualitative and quantitative analysis of data - i.e., marks, predicted marks, record books, and pupil sort	• Teachers summative mark for each marking period based upon the completion of many and varied assigned tasks at an appropriate level of difficulty and standard of mastery • Teachers' marks are task focused and classroom bound, influenced by a host of procedural and contingency rules • Teachers' assumption about the functions of marking are discrepant to the functions ascribed to marks or grades by society

88

Table 4 cont.

Study	Purpose	Foci	Theoretical Orientation	Methods	Findings
4. Yinger (1980)	To describe one teacher's planning decisions	Planning for instructional activities and the use of teaching routines	Interpretive (Participant observation)	Classroom observation, interview, and think aloud techniques; teacher of combined 1-2 grades classroom	• Activities and routines were the most important and most frequent planning concerns of the teacher studied. • Routines were also central tot his teacher's planning behavior such that the author characterized planning as decision-making about the selection, organization, and sequencing of routines.

Moreover, Yinger, like several other researchers (e.g. Zahorick, 1975), found that the basic unit of planning for experienced teachers was activities rather than goals or objectives. Teachers apparently tend to judge the adequacy of their plans in terms of student reaction and involvement rather than achievement and may focus on quantity rather than quality of student participation in evaluating how well an instructional episode or activity is progressing (Clark and Peterson, 1986).

Research on teachers' interactive decision-making informs us that when teachers do focus on student characteristics, decisions about an individual or group of students are made quickly and efficiently (Calderhead, 1983). However, because teachers have a tendency to form their impressions of students early and often on the basis of nonacademic criteria, there is the possibility that teachers may often misjudge learner potential. The Gil (1980) study confirms this result, finding that teachers frequently misdiagnose a student's reading difficulties and use quite different processes for arriving at a case diagnosis. Gil bases his research on the premise that there are so many student in school today who are experiencing difficulty in learning or advancing their reading skills that there is neither the time nor the money to have highly trained specialists for all those in need. Hence, the classroom teacher will need to assume more responsibility for the proper diagnosis and treatment of students with reading problems. While Gil seeks a practical solution to a complex problem, his research also reveals that the fast-paced assessment and decision-making world of the classroom may be overwhelming for the teacher when the assessment demands require more precision.

The study by Whitmer (1983) focuses on an aspect of postinstructional decision-making that has profound impact on the school careers of students. Whitmer investigates the thinking and judgment processes of teachers when determining a

students' achievement for a marking period. Using a multi-method approach to data collection and analyses, Whitmer discovered that teachers hold radically different assumptions about the meaning of marks (grades) than society at large. Teachers marks are task focused and classroom bound. Rather than expected to serve as measures of mastery, predictors of future success, or even motivational devices, student marks are based primarily upon the "neatness and completeness" doctrine. Student work assignments are typically evaluated on how many, how well, and at what level of difficulty a student completed assignments.

Finally, the Stern and Shavelson (1981) study represents a rare overlap in areas of focus of the present study. These researchers were interested in discovering relationships between teacher judgments of student ability, their classroom grouping decisions, and their instructional behavior with the different groups. This study threads through the literatures on ability grouping, classroom pupil evaluation, and teacher decision-making. It is more similar to the present study in this regard. Stern and Shavelson found that teachers group students on the basis of ability even though they have access to many other sources of information about their students. They also found that once grouped, the group and not the individual student became the primary unit for instructional planning and decision-making. And, they found that teacher's instructional plans and behaviors differed markedly between high and low ability groups. These are significant findings which, when considered together, cast a dim light on the school practice of grouping children by ability.

The review of related research on teacher thought processes indicates that the decision contexts teachers face are incredibly complex and demanding. Studies also show that whether their judgments are accurate or not, based on quality information or

not, deliberately reached or not, teachers are crucial decision-makers in the lives of children. They hold the key to changing conditions that will improve or hinder learning for youth while they are in school. A few studies provide some evidence about how teachers use data about learners to inform classroom decisions. In general though, not much is known about how teachers use information that they have gathered on student learning progress. The present inquiry sought to produce data that would be useful in addressing this question.

Chapter Summary

The present study moves away from the emphasis in previous research on teaching which seeks to correlate specific teacher behaviors with student achievement. It shifts the teacher to the foreground of the inquiry as a crucial evaluator and decision-maker. Instructional grouping decisions by teachers are seen as a function of pupil evaluation. A selected review of studies of ability grouping, classroom evaluation, and teacher decision-making was presented to identify some of the existing gaps in knowledge in these related literatures. The review of research suggested a need for:

- increased understanding of the criteria teachers use to make group placements for instruction

- increased understanding of the evaluation methods teachers use to assess student learning progress and whether these methods differ for students in different groups

- increased understanding of the kinds of decisions teachers make as a result of their classroom evaluations of student learning

Three corresponding research questions were framed to extend present knowledge of classroom evaluation environments.

The discussion in this chapter first considered the meaning of evaluation as inferred from the work of various scholars in the field and situated the present study

within interpretive and critical theoretical perspectives. An interpretive perspective was adapted in the present study for its potential for enriching current understanding of the complexity of evaluation at the classroom level. A critical perspective was also adapted because of its potential in identifying school and classroom conditions that promote inequalities among learners. The following chapter presents the research methods used in the study that complement these perspectives.

CHAPTER III

RESEARCH PROCEDURES

The purpose of this chapter is to describe the design of the study. The processes involved in selecting the sample schools and teachers is detailed. Next, the development of the interview instrument and the preliminary testing of the instrument are described. The data base for the present study includes the in-depth interview responses from 23 experienced 5th and 6th grade public school teachers serving 515 children in 10 schools in western Massachusetts. Finally, procedures for collecting and analyzing the data resulting from the interviews are described.

Design

The central purpose of this study was to discover the criteria teachers use to group students for reading instruction, the means they use to evaluate reading progress for the individuals in the various groups, and the ways they use the information they obtain from pupil evaluation. For this study, data came from a qualitative research approach. The intention was not to control the language and substance of what teachers express. Rather, the design tried to maximize opportunities for teachers to say what they wanted to say about pupil evaluation within a broad but thoughtfully considered framework of questions and topics (Patton, 1980; Bogdan and Biklen, 1982). The study discovered teachers' perceptions toward criteria for grouping procedures, for evaluating student progress, and for uses of student evaluation primarily by invitation and conversation rather than by intervention and regulation (after Thiessen and Morehead, 1985). Data produced are descriptive based on teachers' verbal responses to three major research questions and associated sub-questions.

This study is characterized by several qualitative design elements that are patterned after Harrison's study on ability grouping (1989) and Thiessen and Morehead's study of teachers' perceptions of their practices in student evaluation (1985). First, research questions and subquestions serve as flexible guides for data collection. Second, data come from open-ended interviews that address a set of issues to be explored with each respondent but permit the interview to flow in a natural conversational way, with all topics covered by the conclusion of the interview. Third, the qualitative design involves obtaining data in a flexible and spontaneous fashion that are systematic and thorough for each participant.

The following sequence of activities was used to enact the research design:

1.   Written invitations were sent to 18 principals in the Coalition for School Improvement inviting them to participate in the study and asking them to arrange a meeting with their 5th and 6th grade teachers.

2.   Each principal was contacted via telephone to determine his/her decision to participate. Principals from 10 schools decided to be involved in the research study.

3.   The researcher visited each of the participating schools and met with 5th and 6th grade teachers who taught Reading in regular classrooms. The purpose of the research was explained and teachers decided whether they wanted to be involved.

4.   Twenty-four 5th and 6th grade teachers across 10 schools decided to participate. Consent forms explaining the details of the interview procedures and ensuring confidentiality were signed.

5.   Interviews were scheduled with all participating teachers.

6.   Preliminary testing was done with four teachers who volunteered to help refine the interview procedures. These teachers were not included in the main study.

7.   Interviews were conducted with 23 teachers. One teacher from the original group consenting to participate decided not to be involved.

This sequence of activity provides an overview of how the initial stages of the research were conducted. Specific details of the research design are explained in the following sections of the chapter.

Selection of the Sample

Because of the delimitations of the study, the sampling procedures used by the researcher did not actually cull the tremendous range of demographic and geographic characteristics that exist among teachers, students, schools, communities, states and regions in this country. Rather, a more realistic approach was that the data collected could reasonably represent selected pupil evaluation practices of elementary public school teachers of Reading, grades, 5-6, in urban, rural, and semi-rural counties of Western Massachusetts. This group of teachers served as the target population.

With the assistance of leaders in the school-university community the researcher was able to gain access to two (2) groups of teachers that were embedded in the target population. Together, these two groups defined the accessible population of the study. They were:

1.   Elementary public school teachers of Reading, grades 5-6, in core schools of the Coalition for School Improvement. The Coalition is a partnership between the Center for Curriculum Studies at the University of Massachusetts, Amherst and demographically different elementary and secondary schools in Western Massachusetts. For the targeted grades in this study there were seven (7) eligible member schools representing four counties - Berkshire, Franklin, Hampden, and Worcester.

2. Elementary public school teachers of Reading, grades 5-6, in affiliate schools of the Coalition for School Improvement. For the targeted grades in this study there were eleven (11) eligible affiliate schools in three counties - Franklin, Hampden, and Worcester.

Sampling Procedures

The primary unit of analysis of an inquiry into elementary public school teachers' perceptions of their practices in pupil grouping and evaluation is teachers. Any effort to investigate classroom behaviors of teachers must first gain the cooperation of the schools in which teachers work. Hence, a letter was sent to coalition principals explaining the purpose and inviting participation in the study. (See Appendix A)

A target number of ten (10) schools and at least twenty (20) teachers representing demographic diversity was sought for the study sample. Since an important objective of the Coalition is "to conduct collaborative research into problems that persist in school and university settings,"[23] all eligible member schools were sought for inclusion in the sample.

School leaders in five of the seven eligible member schools responded to the initial invitation in the affirmative.[24] These five schools were selected for participation. To round out the school sample, follow-up telephone calls were made to six affiliate schools, selected on the basis of their geographic location and demographic make-up of student population. Principals of five of these schools showed interest and willingness to participate in the study. Thus, there were ten (10) schools - five core and five affiliate, that participated in the main study.

The 10 schools represented a variety of organizational patterns and types of communities. There were eight elementary schools, one middle school, and one high

school. One of the elementary schools had a grade span of pre-K through grade 5, one was K-5; one was pre-K through grade 6; three were K-6; one school had a K-8 span and one school had a 4-6 grade organization; the middle school in the sample had a 4-8 grade span; the high school had a 6-12 grade scheme. Four of the schools were in urban communities, one school was in a suburban community, three were in rural areas, and two schools were classified as being in rural/suburban communities with one school serving students in the district as well as those bussed in from a nearby urban center. In socio-economic terms, two schools were located in affluent communities, two in middle class communities, four in working class neighborhoods with one of these classified as poor (70% of the children receiving free lunches), and two in what is described as diverse or a combination of the above categories. Eight schools had a primarily white racial make-up, with white students accounting for 93% to 100% of the student population. One school had a predominantly hispanic composition of 79% hispanic, 15% white, and 5% black. One school had a somewhat mixed student population of 82% white, 10% black, 6% Hispanic, and 2% Asian. The schools ranged in student body size from 138 to 616. The number of teachers ranged from nine to 42. Schools in the sample typically had two to four teachers in their fifth and sixth grades. Table 5 shows a composite of selected characteristics of the school sample.

To select the teacher sample, each consenting principal was asked via phone conversation to arrange a meeting with the researcher and experienced teachers of Reading in grades fifth and sixth in the building. The purpose of this meeting was to discuss the research, detail procedures, answer questions, and invite teacher participation in the study. A meeting with eligible teachers in each of the ten schools yielded twenty-four (24) interested teachers, each of whom signed written consent forms that described

Table 5

Selected Demographic Characteristics of Member and Affiliate Schools in the Coalition for School Improvement Participating in the Study (N = 10)

School Member	Grades	School Population Students/Teachers		Neighborhood	Socio-Economic Description	Town/City Population
A	K-6	138	10	rural	middle class	1,250
B	K-6	581	26	urban	working class	52,000
C	pre-5	616	25	rural	working class	7,500
D	K-8	303	15	rural/suburban	working class	11,400
E	4-8	568	40	rural/suburban	diverse	7,250
Affiliate						
F	K-6	609	27	urban	affluent	164,650
G	4-6	487	23	urban	working class	44,700
H	6-12	377	42	suburban	affluent	6,000
I	K-5	193	9	urban	middle class	52,000
J	Pre-K-6	252	11	rural	diverse	2,400

Note: Population figures are based on the 1980 U.S. Census

99

how the rights, welfare, and identity of participants would be protected (See Appendix B). Subsequent interviews for each teacher-participant were scheduled at this meeting as well. At the time of the scheduled interview, one teacher declined to be interviewed. Thus, the sample for the study included twenty-three experienced Reading teachers in grades five and six.

There was a fairly even distribution of male and female teachers and of fifth and sixth grade teachers in the study sample. They were a seasoned group of professionals, comparable in their overall years of teaching experience, years teaching Reading, and education levels. Of the 23 teachers comprising the sample, 10 were 6th grade teachers and 13 were 5th grade teachers. There were 13 males and 10 females represented in the study. Sixth grade male teachers averaged: 19 years of overall teaching experience (range 6-32 years), 18 years teaching Reading, and six years teaching grade six. Sixth grade female teachers averaged: 17 years of overall teaching experience (range 12-25 years), and 14 years teaching Reading, and 11 years teaching grade six. Fifth grade male teachers averaged: 17 years of overall teaching experience (range 14-21 years), 16 years teaching Reading, and eight years teaching grade five. Fifth grade female teachers averaged: 20 years overall teaching experience (range 10-32 years), 19 years teaching Reading, and seven years teaching grade five.

In addition, 11 of the 23 teachers (six males and five females) had bachelors degrees, nine (three males and five females) had masters degrees and three (males) had a masters degree plus at least thirty additional graduate credits. Table 6 shows a composite of these variables for each of the teachers in the study.

Table 6

Profile of Teachers Participating in the Study
by Sex, Grade, Teaching Experience, and Educational Level
(N = 23)

	6th Grade					5th Grade			
	TE	TR	GL	HD		TE	TR	GL	HD
Male	12	12	4	B.A.		15	8	8	B.A.+21
	6	6	3	B.A.		14	14	1	M.A.+9
	27	27	19	M.Ed.		21	20	12	B.A.+24
	32	32	15	M.Ed.+30		19	19	15	M.Ed.
	22	22	17	M.Ed.+60		20	20	15	B.S.
	10	7	7	M.Ed.		20	20	5	M.Ed.+30
						11	11	3	B.S.
Totals	109	106	64			120	112	59	
Average	19	18	6			17	16	8	
	N = 6					N = 7			
Female	15	15	15	B.A.		22	22	8	B.A.
	25	22	17	M.A.+9		16	16	5	M.A.
	15	8	8	M.A.		10	7	1	B.A.
	12	12	2	M.Ed.		32	25	19	B.A.+24
						13	13	5	M.Ed.+12
						28	28	2	B.S.+24
Totals	67	57	42			121	111	40	
Average	17	14	11			20	19	7	
	N = 4					N = 6			

Abbreviation Key: TE = Number of Years of Overall Teaching Experience
TR = Number of Years Teaching Reading
GL = Number of Years Teaching at Present Grade Level
HD = Highest Degree Attained

Instrument Development

The researcher has variations to choose from in qualitative interviewing techniques of gathering data. The three basic choices are the unstructured interview, the semistructured interview, or the structured interview. From an interpretive research perspective, corresponding terms might be intuitive, inductive, and deliberative interview processes (Bogdan and Biklen, 1982). Each approach has advantages and disadvantages. The basic difference in the three approaches is the extent to which the questions are determined and standardized before the interview occurs (Patton, 1980). On a continuum ranging from the spontaneous generation of questions in the natural flow of conversation to a set of pre-arranged open-ended questions, Figure 1 below indicates the interview approach used in the present study. The approach is characterized by structured questions, but there is no set sequence that dictates the interactions between the interviewer and the teachers reporting their perceptions.

_____X_____

unstructured	semistructured	highly
interview	interview	structured
		interview
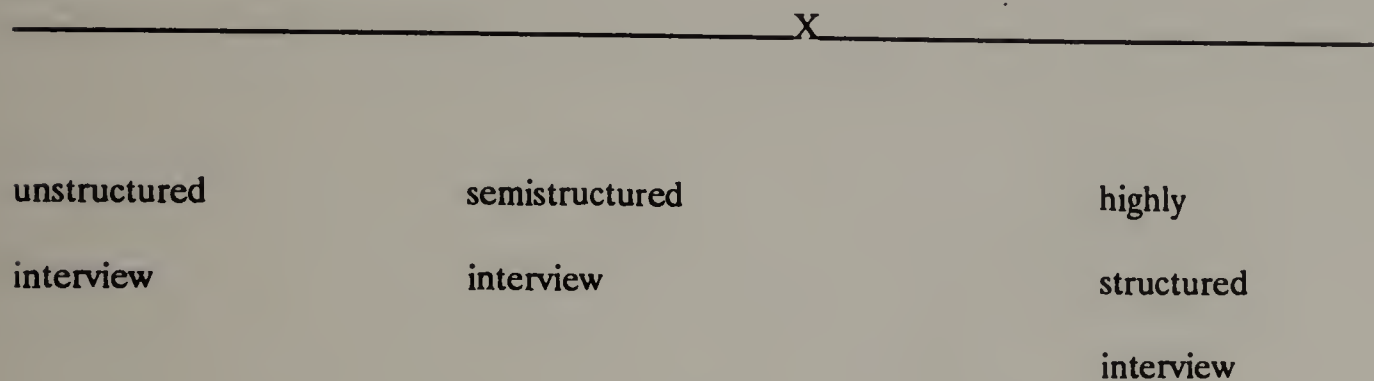
Figure 1

Interview Approach of the Present Study Designated on an
Unstructured - Highly Structured Continuum

As Michael Patton (1980) succinctly states, "the purpose of interviewing is to find out what is in and on someone else's mind." (p. 196) Spradley (1980) also paraphrases this notion when he suggests, "If we want to find out what people know, we must get inside their heads." (p. 10) Finding out what is in and on someone else's mind is not an easily accomplished task. There are many factors that influence and can cause one to report information that is different from what one is actually thinking and feeling. Evaluating students is an uncomfortable task for many teachers. The issues involved in student evaluation are complex. Teacher practices in this domain are elusive and difficult to observe directly. As a result, specific procedures were taken in the planning and conduct of this study in an effort to produce perceptual data that was accurate and honest.

Pilot Study

This present research involved a small pilot study devoted to the refinement of the interview questions and procedures. Interviews were conducted with four intermediate grade teachers of Reading in four elementary schools. First the three main questions were assembled into a written guide:

1.   What criteria do you use to group students for instruction in Reading?

2.   What are the similarities and differences in ways you evaluate the progress your students are making in the Reading groups?

3.   What are the ways that you use information from your evaluation of student learning in Reading?

As a result of this preliminary testing of the interview procedures, useful refinements were made that increased the likelihood of the interview becoming a meaningful way to collect data about the three major research questions. These

refinements can be summarized in five parts: 1) length of the interview; 2) preliminary data to obtain from each participant; 3) related topics and issues to be addressed as part of each main question; 4) interview consistency; and, 5) communication techniques.

Length of the interview. The initial interview took one hour and 30-45 minutes in length. While an abundance of data resulted, teachers agreed that the duration of the interview needed to be shortened if the researcher was to maintain the attention of respondents and gain cooperation of an unfamiliar group of professionals. It was recommended that the interviews last 45 minutes to an hour.

Preliminary data. It became apparent from the preliminary testing of the interviews that selected background information may be useful for understanding the diversity of respondents. It was recommended that information about teaching experience, subject and grade level experience, and education be included in the preliminary information for the interviews. Further it was recommended that additional information be gathered about the classroom context in which ability grouping for reading takes place. A brief description of the classroom organization, alternative grouping patterns, and various approaches to teaching reading are also included in the preliminary information. (See Appendix C for questions, Appendix F for analysis of results.)

Related subquestions. The data collected during preliminary testing resulted in substantive topics that could help in answering the major research questions. These topics were recorded and subquestions prepared to reflect the various suggested topics. Following are the three major research questions and the resulting subquestions.

Research Question 1: What criteria do teachers report that they use to group pupils for instruction in Reading?

- What kinds of assessments do teachers use to assign students to Reading groups within their classroom?

- How comfortable are teachers with the assessment strategies used to assign students to Reading groups within the classroom?

- Of the types of assessments teachers use to make group assignments in Reading, which do they rely on most?

- What goals and expectations do teachers have for their Reading program.

- Are teachers' stated goals and expectations similar or different for each of their Reading groups? How are they similar or different?

Research Question 2: What are teachers' perceptions of the similarities and differences in ways they evaluate pupil learning progress in instructional groups for Reading?

- How many instructional groups do teachers have in their Reading classroom?

- What specific methods do teachers use to evaluate student learning progress in Reading groups?

- What performance criteria do teachers use to judge student work in the Reading curriculum?

- Are teachers' evaluation methods and performance criteria similar or different across the Reading groups?

- Do teachers describe their approach to pupil evaluation in Reading as primarily comparative-referenced or criterion-referenced?

Research Question 3: What are teachers' perceptions of ways they use information from their evaluation of pupil learning progress in Reading?

- What are the ways that teachers use data from pupil evaluation in Reading?

-   How many children have teachers moved from one Reading group to another this year?

-   If movement, in what direction have teachers moved children?

Preliminary testing indicated that it was important to provide opportunities during the interview to ask teachers if they had additional information to add to their responses. Further, it was recommended during preliminary testing procedures that it may be useful for teachers to provide actual samples of evaluation data to more carefully illustrate the various ways they use evaluation data. This recommendation was intended to foster accurate communication between the respondents and the researcher for the collection of data associated with the third research question. In short, related subquestions developed from preliminary testing helped the investigator answer the major research questions and helped teachers to more clearly understand them.

Interview consistency. Perhaps most important, pilot study interviews reinforced the importance of standardizing procedures for recording data. Four important steps were utilized in the main study to record responses accurately and completely.

First, it was recommended that a cassette recorder be used to record all data resulting from the interview. In case of technical difficulties with the primary recorder, a backup micro-cassette recorder with matching micro-tape in place and ready for immediate use would be available. Second, it was suggested that the researcher note important responses to questions that may assist in the interpretation of information that was tape-recorded. In this way, the investigator could listen attentively to the respondent while at the same time writing down main points in the margins of the interview guide. Also, it was recommended that sufficient space be left for each open-ended question and subquestion to permit quick and efficient note taking. Through

this procedure, data analysis was facilitated as the researcher collected responses to interview questions. Third, it was recommended that all tape recorded data be transcribed for each interview. These transcriptions should be read carefully by reviewers for completeness and accuracy of communication. Fourth, the preliminary testing showed that it may be helpful to review the responses on the tape recorder and the notes to distill specific substance that is directly related to answering each of the research questions. This review should include judges who analyze the interview responses to confirm the association between the data and the answers to the research questions.

In short, the above suggestions gleaned from the preliminary testing of the interview procedures helped to improve the substance and design of data collection and better ensure consistency in the interviews used to collect data for answering the research questions that guided the study.

Communication during the interview. The purpose of an effective interview is to obtain accurate and adequate responses for each question. Carol Weiss (1975) has suggested that both the respondent and the interviewer can have predispositions going into an interview that can lead to errors or a difference between the answer that is given by a respondent and the true answer. Thus, any strategies that can be shown to reduce this difference can be said to increase the accuracy of the interview.

At least four communication techniques were suggested so that interview errors were minimized. First, wording of questions was perfected so that they were clear and understandable to teachers. Second, those questions judged to be misleading were identified and eliminated from the interview. Third, although affiliation with the Coalition for School Improvement helped the interviewer gain the cooperation of

teachers, it was still necessary to establish personal rapport with each teacher who participated in the study. It was suggested that a special effort be made at the outset of the interview to thank the teachers for their willingness to help with the research. Fourth, as a result of the preliminary testing, it was decided that one person should conduct all interviews. This was intended to ensure that each participant was having the same interview.

In sum, preliminary testing helped the researcher reduce the length of the time of the interview, include preliminary data about each participant, develop subquestions for each major research question, create ways to ensure consistency in the collection of data, and communicate effectively with respondents. The final section of Chapter III summarizes the research procedures used to collect, report, and to analyze the data.

Data Collection

The data base for the present study includes the in-depth interview responses of 23 teachers in 10 public schools. All suggestions resulting from preliminary testing of interview procedures were utilized in the data collection. A one hour interview was conducted with each of 23 teachers who participated in the study. The investigator conducted all 23 interviews using cassette recorder and an interview guide containing: the preliminary information sheet, main research questions and related subquestions. The data collection period lasted from March 20, 1989 to May 15, 1989 and required more than 2,500 miles of travel to the 10 participating schools.

To facilitate the collection of information from teachers, and to protect the anonymity of each participant three complementary data collection procedures were used. First, each interview was recorded on audio-cassette. Ninety minute tapes were used to allow sufficient time for interview. Each tape was given a 5-digit numerical

code in accord with the school, teacher, and grade being interviewed. Transcribed interview data retained this same code number. Only the researcher had access to code identification numbers and particular teachers. Hence, no teacher names were ever used, orally or in writing, to identify a teacher with his/her responses. Second, each interview guide had a corresponding code assigned to it in the upper right hand corner of the first page. And third, a special copy of the three main questions was prepared and presented to each teacher in the study at the beginning of the interview session in a sealed envelope. The purpose of this procedure was to allow time for respondents to reflect further upon their responses and mail any additional information to the investigator within ten days after the interview. The teacher's code was lightly pencilled on the back flap of the pre-stamped return envelope to keep data organized.

Data Analysis

In order to analyze teacher responses systematically, procedures developed by James Spradley (1979, 1980) were adapted to guide the analysis of data for the present study. In accordance with Spradley's procedures, teacher comments were transcribed, read and re-read carefully to determine a set of domains for classifying their content.[25] This is referred to as *domain analysis*. For example, three domains emerged from an analysis of responses to the question about criteria used to assign students to Reading groups: 1) grouping criteria used before instruction begins; 2) grouping criteria used during the first six weeks of school; and, 3) grouping criteria used throughout the remainder of the year.

Within each domain, responses were then subcategorized in terms of their semantic relationship to the domain. This is referred to by Spradley as *taxonomic analysis*. In the domain of "grouping criteria used before instruction begins," for

example, four categories emerged, 1) Recommendations from other professional staff; 2) Standardized test results; 3) Basal reading programs; 4) Teacher's own pre-assessment strategies. Further, each subcategory was analyzed to determine the characteristics which distinguished it from other subcategories. This procedure is called *componential analysis*. For example, within the domain of "grouping criteria used before instruction begins," and subcategory "standardized test results," for example, comments were distinguishable by whether they referred to standardized achievement tests, or state-mandated tests, or tests that accompany basal reading programs. Then, from these contrasting dimensions, patterns or themes that appear consistently throughout or across the set of data were analyzed. Spradley refers to this procedure as *thematic analysis*.

Finally, teacher-interview transcripts and related field notes organized in order of occurrence, were analyzed. Whenever field notes seemed relevant to a particular domain or category, a notation was made in the transcript margin. Comments from field notes were double coded if they contained information pertinent to more than one domain or category. In all, respondents made 496 comments from transcripts and field notes for Research Question 1 and its subquestions, 362 comments for Research Question 2 and its subquestions, and 516 comments for Research Question 3 and its subquestions. Interview and field notes that were not classifiable were eliminated from the data analysis.

Chapter Summary

The research design for the present study employed qualitative research methods to describe the perceptions of experienced 5th and 6th grade teachers of Reading toward three major research questions. Data were collected from 23 in-depth interviews with teachers representing diverse core and affiliate schools of the Coalition for School

Improvement. The researcher conducted a preliminary testing of the interview procedures for the purpose of establishing safeguards against bias and flaws in the design, and to facilitate the process of deliberate yet flexible inquiry. Three major research questions and seventeen follow-up questions guided the open-ended, semistructured interviews.

The researcher examined interview transcripts, and field notes for emerging patterns descriptive of grouping criteria, classroom evaluation methods, and teacher use of pupil evaluation. Sample size was determined by the number of eligible teachers in coalition schools who volunteered to participate in the study. This number was close to 100% (23/24). The primary data for the study were the teachers' own words, supplemented in some cases by official documents (e.g., classroom lists of students, report cards, teacher record book) and other artifacts (e.g., samples of student work, student folders, and culminating projects). The researcher employed a systematic process of classifying and coding data in relation to identified patterns in order to organize data for further analysis and present evidence to support implications of the research questions.

CHAPTER IV

ANALYSIS AND INTERPRETATION OF FINDINGS

This study describes how selected 5th and 6th grade teachers of Reading perceive that they group students for instruction, evaluate their progress, and utilize evaluation data. The chapter is divided into three sections which correspond to the major research questions that guided the investigation.

Research Question 1:

What Criteria do Teachers Use to

Group Students for Instruction in Reading?

Interview data from sample teachers reveal that they rely on two major sources of information - outside sources and their own - to form instructional groups within their reading classrooms. Three domains of grouping criteria emerged from the database for Research Question 1: - 1) grouping criteria used before instruction begins; 2) grouping criteria used during the first six weeks of school; and, 3) grouping criteria used throughout the remainder of the school year.

Grouping Criteria Before Instruction Begins

Although teachers are usually given a list of students already pre- assigned to ability-based reading groups, four categories emerged from analysis of the data for grouping criteria used by teachers to make initial group placements before instruction for the school year gets underway. The categories are: 1) Recommendations from other professional staff; 2) Standardized test results; 3) Progress in basal reading programs; and, 4) Teachers' own pre-assessment strategies. Recommendations from other professional staff was by far the most frequently cited criterion for placing students into Reading groups. Twenty-one out of 23 teachers reported that the recommendation of

the student's previous teacher was a major source of information in making initial classroom grouping assignments in reading. Other professional staff playing key roles in placement decisions included Reading Specialists and the Guidance Counselors. Four teachers also reported that they consulted with the regular and special education teachers to gain a more in-depth perspective of one or more students.

Students' scores on reading portions of standardized tests such as the California Achievement Test (CAT), California Test of Basic Skills (CTBS), and the Metropolitan Achievement Test (MAT) as well as results of state mandated tests of reading achievement were reported to be important criteria. Twelve of the teachers perceived test scores to be an important criteria used by other professionals to recommend student placement decisions. Two teachers said they were not sure of criteria used by others to recommend placement of children in reading groups. This knowledge gap seemed to stem more from astonishment at the number of misplaced students in their view than from ignorance of the actual process. Moreover, the teachers in the study reported that they did not personally consider test scores to be an important criterion for determining their within-class reading groups. Five teachers said they reviewed the comments, anecdotal records and other relevant information in a student's cumulative folder and considered these data when making initial placements for reading groups.

A third category was the use of the extent to which students had completed basal readers and worksheets. Students in schools using basal programs were placed in groups to resume at the level attained in the basal series for the previous year. Also, formal pre-tests that accompany the basal series were used in a complementary way to determine initial reading groups.

Fourth, three teachers identified several of their own strategies - the use of informal observation and reading inventories as criteria for initial groupings. Teachers reported that they listened to students read aloud and asked them questions about their reading habits. Their oral reading performance and past reading practices were then used to place students into initial groups.

In sum, the recommendations of previous teachers and the results of standardized testing were the two leading placement criteria reported by teachers to be used at this initial stage. Individual comments of teachers were compared, contrasted, and analyzed for discussion by adapting procedures described by Spradley (1980), Bogdan and Biklen (1982), and Goetz and LeCompte (1984). Several comments that are indicative of responses that support this finding for grouping criteria at the beginning of the year follow:

> The first criteria that I use is teacher placement from last year and concurrent with that, achievement scores and other testing that may be available in the student's folder. That is the background information that I start with day 1.

> It's already been grouped before we get them. The Guidance Counselor groups them according to previous teacher recommendations and CAT Tests, so they're grouped into three groups of what we call: high 6th grade and above; high 5th grade and low 6th grade level in the second group; and then below that in the third group. [Each teacher gets one group].

> In the beginning of the year we're handed a class list of reading groups that is done by the previous teachers. Then what I do is go over the list, and go through files. Then, generally in the beginning of the year, I do pretty much what is recommended, and then I work from there. Usually the third week or so, I make some changes.

> Early on in the year, we take it from the previous year's teacher. They send us a list of the way they had them grouped, and that's the way we would, of course, start out. They usually come down to us with three groups.

114

In the beginning of the year, we use the teachers from last year suggestions. They give us groups, lists I should say, dividing the children into high, medium, and low. If we have a new child come in, we have a Reading Resource teacher who gives him a test to place him at reading level.

I base a lot of my placement on the level attained from the previous year. With the child's school folder will come a reading skills chart, all the magazines and series that he has passed. Any areas that he has had difficulty with are outlined and his present reading level will be indicated on that chart. At the end of the previous year, the city reading teacher will evaluate, with the classroom teacher, all the children with indications whether they can be moved up, accelerated beyond the level that they are presently working, or if they should go back and redo some of the skills that they have learned. That's the basis with which I start in my Reading program in September.

Thus, teachers seem to share a casual tension toward the process that results in the predetermined class lists (or groups). They officially accept the preliminary grouping ritual while unofficially harboring some skepticism toward it. From a theoretical perspective, teachers seem to be saying that at this point in the year they may be most cognizant of the "errors" that cumulate for the individual student from group instruction. In <u>Human Characteristics and School Learning</u>, Benjamin Bloom (1976) argued:

> Group instructional procedures employed with individual students who vary in many characteristics must produce variations in the accomplishment of a learning task - both in the level of achievement of the task and the rate at which it is accomplished.

Meanwhile at this pre-instructional phase of the year, teachers were preparing to usher learners through yet another round of grouping procedures further compounding favorable learning conditions for some and less favorable learning conditions for others. The idea of accepting (but not really accepting) the recommendations of peers as a criterion for group placement of students is an interesting interpretation of the meaning perspectives of teachers. Teachers did not express dissonance about the possibility that

115

their own end-of- year recommendations may be similarly questioned by the next grade teacher.

Grouping Criteria During the First Six Weeks of School

During the first six weeks of the new school year, all teachers reported that they spend time getting to know the children and employ formative evaluation strategies to make adjustments in their pre-assigned reading groups. Categories were developed from data analysis to describe the various grouping criteria teachers use to launch their reading curricula at this crucial intermediate time.

Four categories of grouping criteria predominate: 1) Basal materials; 2) Oral reading 3) Teacher observations; and, 4) Teacher questions. The materials that accompany the basal series adopted by the school usually include the basal text, workbooks, skill sheets, vocabulary and comprehension exercises and unit or section tests. These materials are claimed to be levelled in accord with discrete, pre-determined skills deemed essential for mastery by a particular age or stage in the learner's development. In schools where basal systems are used, students' skill levels and progress in the published materials are the major criteria used for determining in-class and between-class group assignments.

Fourteen teachers interviewed reported student fluency in oral reading activities as the most frequent way of recasting reading groups in the classroom. Oral reading includes listening to students read in the whole class, small groups, or one-on-one. It also includes the teacher reading aloud to students for their enjoyment and observing their listening skills.

Observations of the learning rate and social adjustment of the child were reported by nearly half the teachers ($N=11$) as common criteria for placement of

students in reading groups during the first month or so of class. Teachers said they closely monitored the academic behaviors as well as the social behavior of their pupils to determine how well students were performing their daily work, the interest and effort being put forth, and how well each child was fitting into their reading group as a whole.

In addition to these four major criteria for formation of within-class ability groups early in the year, four other categories, tailored more to the teacher's individual philosophy, were also identified. First, four teachers using a literature-based approach to classroom reading instruction reported the use of novels to observe students' breadth of prior reading experience, comfort levels with particular reading material, and book selection as criteria for placing students into reading groups.

Second, writing skills were seen as crucial to evaluation in Reading at the intermediate grades and beyond. Three teachers noted that there was a strong positive correlation, in their view, between good reading and good writing. Ample opportunity for students to write about what they had read and to read what they had written were provided in these classrooms.

Third, teacher-made or teacher-adapted materials were used to assess Reading skills of learners to generate criteria that could be used for grouping. Seven teachers reported that they wrote their own comprehension questions for students and utilized their own evaluation materials developed over years of teaching experience to be able to get an idea of learners' reading capabilities.

Fourth, six teachers emphasized the importance of getting to know the child before making any permanent group placements of children of similar reading ability. The following quotes are indicative of how these teachers felt about the importance of getting to know the children:

Normally, initially at the beginning of the year, I teach the whole class for approximately a month until I get to know the students. I also talk with previous year's teacher, but you need that month to get to know the students.

By the fourth week, I'm settled. That doesn't mean that in another few weeks I wouldn't switch somebody. But as a rule, it's [my groups] pretty solid...this is also the time when I have begun to know my kids. Initially, in the beginning, you don't always see a kid for what he really is. Once they become more comfortable with you, they're more apt to show you their strengths and weaknesses. That is a big help for me.

When they come into my class...they are in groups from last year, so I maintain those groups for awhile and continue them in work that the group would be doing based on the unit they are supposed to be entering and kept tabs on whether the work seems challenging for them or not, their attitude about what they are doing, if they finish the stories in a very rapid way and they can relate to me the basic material of the story, whether they are having difficulty with the story after having extreme difficulty with the workbook pages and the vocabulary, and after awhile, I start to get a picture of my own idea of the child's learning style for one thing and also their competence level and how they can handle the level to which they have been assigned...So it's also trying to get to know the child personally over a month and I spend the first four weeks talking with the parents regularly once a week, talking on the phone about these very issues, letting them know what I'm up to and what I have in mind. I talk to all the parents of the kids on the phone once a week and then it tapers off over the year whether we need to talk or don't need to talk. That gives me a picture of the whole child, the home life, reading habits and in talking with the parents, I find out how much time they have to give to the kids and it gives me a much better picture of why they are who they are.

And fifth, in a kind of miscellaneous category, two teachers said that they convened individual meetings with students to discuss reading strengths and weaknesses in order to determine proper placement. In one urban school, the Reading Specialist ran further assessments on children in lower ability groups and gathered these children together for small group instruction. Timed tests, quizzes, and levelled materials from Science Research Associates (SRA) reading program were used systematically by one teacher. Another teacher experimented with mixed ability groups early on putting

"highs" with "lows" and vice versa for extra motivation of "lows". And one teacher, assigned a high homogeneous reading class, assumed that the pre-groupings were valid and did not further assess the students' reading ability. Although fewer teachers reported that they used the latter five categories of criteria for placement of students in reading groups, the intensity of commitment that was expressed toward these criteria suggest the possibility that they are as powerful determinants of student group placement within the classroom as those that are mentioned more often by the selected teachers.

In sum, the perceptions of teachers gathered through interviews suggest that the first weeks of school are characterized by continuous consideration of student reading performance and prior reading behaviors. Also, consideration is given to using the resulting information as criteria for solidifying classroom groups for reading. The data indicate that the predominate pattern for grouping students for reading is by ability. Thus, although additional information about students is collected during this period of time that could be used to place students into groups for Reading instruction, the use of criteria for grouping by ability prevails.

Several teacher comments reflected the range of responses representing this domain of analysis:

> During the first two weeks of school, the class is reading as a whole, therefore, no grouping takes place. During that time, there is observation, both in terms of large class settings, oral reading, and answering questions. At times, for particular students, there is individual follow-up if I am concerned about a particular area. On the basis of those criteria, reading group placements are assigned...

> I would also include what is of interest on the part of the students, therefore, some of the groups that are formed have student input. For example: With the last series of novels (historical fiction) I had three choices. We told the students about the three choices and then asked

for their input in terms of their own interests for particular novels. I would use that as part of my information in terms of forming groups.

My curriculum is based on analysis and reading as a very analytical kind of reading so I need to be sure they're in reading groups that allow that, as far as being a similar level so I do different evaluative oral exercises on oral reading, finding and explaining reasons for actions of characters and the intent of authors, and also their own reactions and responses to what they are reading.

There are a couple of my own tests that I use...not just written...I will sit individually with students and listen to them read orally and ask them questions orally.

I have a fifth grade and they are grouped according to their test scores. When I see a discrepancy, I weed them out.

...I then make my own evaluation...I take about, usually ten days of review for each of the groups...we're generally assigned three groups. For each of the groups I do about ten days of review. During that review process, I re-evaluate the children myself using my own criteria...not necessarily using the present basal reader, the Houghton Mifflin materials. If I see the children need reinforcement in, say, dividing words into syllables or alphabetizing, I'll give them papers on that. If I see a child that seems to be reading beyond where he has been placed, I will re-evaluate him with reading materials that would indicate whether he should be accelerated or skipped to a higher level.

Grouping Criteria Used Throughout the Remainder of the School Year

Twenty-two out of 23 teachers interviewed reported that they continued to make decisions regarding a change in group placement for some students. In classrooms where basal systems guide the reading curriculum, basal materials continue to be the primary means for gathering data about criteria for regrouping students. In classrooms where reading a variety of literature is fostered and writing is a central activity, a broader range of grouping criteria continues to be used. The substance of these criteria seems to reflect the individual preferences of teachers.

The most oft-cited criteria used by teachers through out the year can by described in three main categories. First, teachers in basal-based classrooms reported

that they relied heavily on the following criteria: 1) read-aloud performance; 2) skills performance in seatwork; and, 3) basal test performance.

Second, teachers who orchestrated literature-based classrooms or who practiced autonomy within a traditional approach, reported that they observed students' interest, selections of novels, and other reading material in the classroom to judge the need for changes in group placement. Teachers in these classrooms favored their own criteria for continuous placement of students into reading groups.

Third, several teachers identified their experience in judging student performance, knowledge of learner potential, conferencing with individual students and discussions with colleagues about individual reading strengths and weaknesses, formal and informal questioning techniques to assess comprehension, and student self-evaluation of their own reading abilities as substantive criteria used throughout the year to determine pupil placement in the prevailing ability groups for reading. Two teachers reported that a school-wide inventory or checklist of reading skills was used to chart student reading progress throughout the year and was used as a basis for making changes in student's group placement.

In sum, analysis of interview data for Research Question 1 suggests that there are three periods of the school year during which teachers are applying various evaluative criteria to form and reform instructional groups for their reading classrooms. The recommendations of previous teachers and the results of standardized testing were perceived by teachers to be leading criteria used to place students into initial classroom groups. During the first six weeks of the new school year, teachers say they generally spend an extraordinary amount of time developing rapport with students, observing their oral reading skills, asking questions, assessing prior reading habits, and using novels or

basal materials to solidify classroom group placements. As the routine of the school year settles in, teachers report that they continue to monitor students' read aloud skills. However, criteria for further shifts in classroom groups during this dominant period are related to whether the reading curriculum is based on individual novel selection or a levelled basal program. Although teachers felt that students were often misplaced by reading ability at the beginning of the year, they did not seem to know how to reconcile this problem in their own within-class grouping arrangements. Throughout the year, they continued to collect a wide a range of information about the social, academic, physical, emotional, and psychological characteristics of their learners; however, the system of ability grouping was perpetuated in their classrooms.

Research Question 2:

What are the Similarities and Differences in the Way Teachers

Evaluate the Progress their Students are Making in the

Reading Groups?

Interview data from teachers suggest that the nature of the reading curriculum influences the way they evaluate the learning progress of their students. Two domains of evaluation methods emerged from the data base for Research Question 2: - 1) evaluation methods in basal classrooms; and, 2) evaluation methods in literature-based classrooms.

Evaluation Methods in Basal Classrooms

In basal classrooms, teachers use formative and summative methods to evaluate the learning progress of their students. Formative methods can be described in seven categories: 1) Observation; 2) Comprehension questions; 3) Skill sheets; 4) Vocabulary

exercises; 5) Supplementary levelled materials; 6) Writing assignments; and, 7) Teacher-made materials. Each of these formative methods are detailed below.

First, eight teachers report the use of observation as a frequent method of evaluation. This includes observation of reading aloud skills, i.e. verbal fluency, enunciation, inflection; observation of social behaviors in group work and during seatwork; observation of physical behaviors such as body position, nervousness, tapping, attentiveness as reflected in facial expressions and eye movement; observation of attitudinal factors such as effort, use of "free time"; and, observation of academic behaviors such as how rapidly a student grasped a skill or concept introduced in the basal curriculum, the kinds of questions a student asks or comments made during oral discussions, the content of his/her writing, and how carefully a student completes assigned work.

Second, eight teachers often are asking comprehension questions of students to evaluate this important element of able reading. Teachers use comprehension questions that accompany basal texts and the story for the day as well as their own questions from other materials used in the classroom including tradebooks, youth magazines such as Reader's Digest, Scholastic News, Weekly Reader, and the like. Two teachers said they employed random questioning techniques, calling on volunteers as well as non-volunteers in small group and whole class settings to assess comprehension of recently read materials.

Third, 10 teachers report that worksheets and workbook exercises account for a significant portion of pupil evaluation in the basal classroom. Worksheets that evaluate discrete skills such as cause and effect, main idea, word referent are often assigned as seatwork following small group reading. Complementary workbook exercises on skills

introduced in the story are often corrected together as a whole class "so that the student knows why he/she missed a question".

Fourth, seven teachers report the use of frequent vocabulary exercises that accompany the basal text as well as one they have designed to compliment a particular tradebook or novel that students are reading as means for evaluating student progress.

Fifth, five teachers in basal-based classrooms in several schools reported using supplementary methods and materials that were levelled and self-paced for evaluation of student reading. These included Sustained Silent Reading (SSR), SRA's which include pre-tests to determine initial level, "Triple Takes", Reader Digest supplements, and other levelled materials.

Sixth, four teachers in this domain said that writing was an important emphasis and tool for evaluation in their reading classes. A variety of writing assignments ranging from basic sentence and paragraph construction to lengthier, more creative assignments such as students' skits, plays, commercials, dioramas, characterizations, and research presentations were used to evaluate student progress in Reading.

Seventh, four teachers reported designing or adapting their own materials on occasion for use in evaluating students' comprehension, vocabulary, and skill development for Reading. Three teachers also reported developing periodic tests and quizzes to assess a small cluster of skills or to assess comprehension and vocabulary.

In sum, these seven categories describe the variety of formative methods that teachers in basal classrooms use to evaluate pupil learning progress in ability groups for Reading. Sample comments are reported below to enrich the reader's understanding of teacher perceptions of the evaluation methods.

> Sometimes, what I see a kid doing in the classroom helps me to mark
> them too. You might have a kid that's doing the reading and he's very

good orally, and he's just not someone who likes to sit down and answer questions about a story he's read. He does poorly on the written work, but he really does understand what he's reading or what he's doing. I do take that into consideration.

Basically, it's all the work in their workbooks. At the end of each magazine, there's a test.

Their comprehension of the story, their skills on work sheets and workbook and on the test at the end of each unit to make sure they have passed in each area and if they don't, then you have to go back and work on that.

No formal test. Like I said, it's just papers that they pass in. I don't give them a formal test. It's just the papers that they produce from their comprehension questions, vocabulary work, supplemental material I use...it's kind of old material, but it's Burnell Locke material I really do like, especially with my low group.

I use the Basal text comprehension questions...I use vocabulary...For instance, let's say I assigned eight or 10 words from a story. The kids look up the word, put it in the sentence, and I'll ask around in the group " What does this mean? What does that mean? So, I use that as a means of evaluation...Comprehension questions, I weigh pretty heavily...I also have a reading kit that I use called "Reading For Understanding"...there is a pre-test given to put them into a level, or where they would begin in the kit. Like a diagnostic test...Also, I use the Weekly Reader...I will usually go over it with the kids. I'll have them read it and then we usually go over it together. I'll plan in what I think are important vocabulary words. A lot of the times I'll let them point out what they think are important vocabulary words. They usually have a section about vocabulary...it's usually matching...and the kids will do those questions. And, I usually grade those.

Primarily, I use worksheets and workbooks with the section test. Usually we do two pages from the workbook and I follow it up with worksheets. We do the workbooks together. Then I pass out the worksheets and they try to do them on their own...Also, what I usually do is every month I take one particular paper. I keep it. By June I see - like, suppose, I had ten questions in September and the kid got maybe four right and his answers were completely out of it. By June maybe this kid has six right answers. I can see some growth anyway. I compare papers like that. And towards the other way I may have a kid who got eight out of ten and by June only six out of ten. So it isn't all positive all the time...The section test comes every quarter, basically eight to ten weeks. Then in June we're going to have section book test on all our subjects which gives me a total assessment on how the kids are doing.

As far as skills go, it's a matter of assigning them workbook sheets or other related vocabulary or comprehension work and checking their answers and seeing if they understand the skill involved or have comprehended the story or understand the meaning of the words at the skill level. The other levels are enthusiasm for reading, being able to answer questions spontaneously and get involved in a discussion about a story after having read it, being able to project ideas from the story out into the world and the children's own experience, and that's some more, I guess, critical thinking area and often times I will have them stop in the middle of the story and write their own ending and then compare their ending with the ending of the story and that lets me know how involved they are with the story as well and how much they identify with it and get into it themselves, actually become a part of the story.

Comprehension questions...They can be from the book or they can be my own. Scholastic News has their own set of questions...sometimes I use those. Sometimes, I make up my own. It can be paragraph writing where they do a paragraph on a particular subject or story they had read, or an item they had read in a newspaper...that I would evaluate and give them a mark for. They do vocabulary work every week that's graded...the words, the pronunciation, the meaning, using them in sentences correctly...Once in a great while, I give them a test...an actual printed test...Sometimes from the text...Maybe twice a year...They have workbook pages that are done. I correct them with them...they correct it...we do it together...So, basically, it's observation, marks for different assignments, and daily class work.

Three categories of summative evaluation methods emerged in the data for basal classrooms. These are 1) Section tests; 2) Writing assignments; and, 3) Reading skills checklists.

Clearly, the most oft-used summative evaluation method reported by teachers was the section test. Eight teachers said they used this kind of instrument to evaluate the progress a student was making in reading. These tests accompanied the basal series and were used to assess students' achievement at the end of the units of their text. These tests typically included items designed to measure comprehension, skill work, and vocabulary development. One teacher provided a useful explanation of the section test as a summative evaluation method:

We use a Basal reader, the Houghton Mifflin series. The Houghton Mifflin series has a standardized test at the end of each section...each magazine of the book...and we are required to give the test and grade it. It has to be submitted to the city reading teacher and she actually determines with the classroom teacher whether the children are going to continue progressing in that book or have to be held back to redo the skills...Each one of those books is divided into four sections. So, I would say every six to eight weeks you're going to finish a section. That's a required evaluation.

Second, four teachers stated that conventional and creatively adapted book report assignments were given to students to check their understanding and critical interpretation of what story or novel they had read. For these assignments, students often were required to write book reports which described what their favorite part of the book was and why they would recommend it to a friend. Other written assignments were described by teachers as summative evaluation methods, including students' writing their own "book".

Third, two teachers reported that they used a Reading Skills Checklist to evaluate student progress and to track it over the course of the current year. Such checklists included general indices pertaining to: attitude toward reading, use of textual clues, (reading) selection skills, reading strategies and comprehension, oral reading, written work, and skill in using reference materials.

Finally, in addition to the evaluation methods described above, some teachers pointed out that it was indeed difficult to explain the way that they went about evaluating student progress. One teacher, for example said:

I have a million evaluation methods...I don't do any one thing on a regular basis.

Another teacher indicated that intuition and spontaneity play a part in pupil evaluation:

It's strictly off the cuff. The only time I'll make notes on somebody is if I notice something very specific...

Evaluation Methods in Literature-Based Classrooms

In literature or novel-based classrooms, teachers also use formative and summative methods to evaluate the learning progress of their students. Some of the emergent categories of evaluation correspond to those identified for teachers using basal approaches. The formative evaluation methods for these teachers can be described in five categories: 1) Observation; 2) Discussion; 3) Writing assignments; 4) Teacher-made materials; and, 5) Individual conferences. Two common summative evaluation methods were described: 1) Creative writing projects and 2) Reader Response activities. Each of these methods is briefly described in turn.

As in basal classrooms, teachers in literature-based classrooms reported using observation as a frequent means of evaluation of student learning. Observation refers to listening to oral reading, detecting skill strengths and weaknesses of higher and lower grouped students, and being sensitive to students' social maturity and "readiness" to become personally involved in reading. Seven teachers in literature-based classrooms reported using this method of evaluation. Discussions refer to whole class exchanges of ideas and small group conversations about the novel being read; including details about setting, characters, plot, conflict, author intent, and evaluative questions that ask students to give their personal judgment of the book. Discussions also refer to teacher-student elaborations of particular skills that are being emphasized during the marking period. Five teachers said they used discussions as a means for evaluating student learning progress in reading.

Written assignments as a means of formative evaluation can be short answer work to check recall of information or facts contained in the novel, teacher-designed comprehension questions, in-class assignments to reinforce isolated skills such as cause and effect or drawing conclusions, and lengthier responses to essay questions to develop critical thinking. Eight teachers reported that written assignments were used as an important means of student evaluation. In addition, in literature-based classrooms teachers seemed to often make or create their own materials and assignments to evaluate how well students respond to the books they are reading. These include question sheets as well as pop quizzes and tests on material covered in the literature piece. Five teachers said they made their own materials and tests to evaluate student learning. And, three teachers emphasized the use of individual conferences with students to ask questions and review progress with students who made individual book selections. Conferences were also a way for teachers to touch base with students, allowing them to share what was going on in their lives in and away from school as well as to discuss specific learning needs.

One distinguishing feature of the evaluation methods of three teachers using literature approaches was the proliferation of creative projects used as summative measures of student reading progress. These were primarily writing projects, but demonstrations in the arts and other subjects were often integrated as students would, for example, design sets for plays, make illustrations and three-dimensional renderings in the form of dioramas, create their own TV guides, advertisements, write their own short stories, poems, plays, mythology, and mini-novels, and perform pantomimes. In addition, two teachers reported the use of a Reader Response activity designed to encourage students to think critically about their reading.

Teachers in literature-based classrooms express in their own words how they apply these evaluation methods in practice:

The main thing that I evaluate is what they write about our discussions of our Reading...when they write at the beginning of the year, I don't mark it all in red, correcting every mistake. I tell them I'm looking for their ideas. I'm looking for them to say as much as they can with the time that they have, in the best way that they can...We read a book together. Everybody gets a copy. We have an assignment, we write about it, we talk about it, we do plays, art projects, and all different activities...One of the essay questions was "Describe how Caddy Woodlawn changed throughout the course of the book. Use at least three examples from the book." So, they had to describe how she was a tomboy at the beginning of the story and had a lot of freedom, and how in the 1860's she had certain expectations for becoming a woman...how she changed and how her father supported her life and encouraged her to get more exercise with the boys, and how he helped her make the transition to become a young woman in the 1860's...I would say that test work is fairly similar at the end. The kinds of questions that I ask have things in common, but there's no special format.

Observation, tests, quizzes, just the assignments I give. Are they able to do the skill. If we are talking about cause and effect and I give them a series of things that happened in the novel and asked them what caused those things, are they able to transfer the skill of cause and effect to the novel they are reading...With quizzes, I am referring more to vocabulary as far as formal quizzes, the informal are just the kinds of questions we are asking and making notations of students who are really not able to answer, who seem to have read it poorly or have missed points. Using miming or doing activity with the word, asking for a synonym, so that through the use of it, I am observing whether they have knowledge of that or not. At the end of the week we do a quiz on the vocabulary...If I have taught something, for instance, the skill of cause and effect and I would like to apply it in their reading. If I mentioned five incidences that happened in a novel, would they be able to get those five causes and draw that conclusion...When they are reading out loud and they mis-pronounce a vocabulary word that they have had all week and still don't know how to pronounce it, or if we are doing pantomiming, did they have trouble with picking which word or acting it out. For example, if the word was to be "peer" or "staring", would they be able to do that or could they not.

As far as the actual evaluation of how they're doing, it depends largely on their written work and how well they elaborate on ideas, points they make, how they use examples, how they explain from whence their reasoning comes. The written work is not short answer work, it's all essays, all explanations of figurative language, of characterizations, and

things to do with how the author has written in the plot. It's very similar in group discussions, and that's part of the evaluation as well. I just look at how well they substantiate points and with many students I have to juggle the ones who express themselves less well or less willingly in a discussion...

We always have at least a ten minute discussion at the beginning of the class. If the discussions begin to deteriorate so that I can feel they're not reading, then I will do pop quizzes or surprise quizzes, as the kids call them. It's related to the reading done the previous day. They also do some kind of a...I hate to use the word "book report" because it's not a book report...According to their own style, some boys and girls will come in dressed as the character in the book and do a scene. Some boys and girls will prefer to do a mobile of the important events in the story. Some of the boys and girls will do what they like...a very concise, written report. I give them several suggestions. Some will do a diary if it is appropriate to the story. Some do dioramas.

This is Rebecca [not the child's name] at the beginning of the year (visual)...She was reading a book called "The Witches" at the beginning of the year and I asked them to do a Reader Response to talk about what they thought about the story. I have a list of things..the Start Ups, I call them...and she said she thought the first chapter was good. Very general. Then I said "If you have any questions...why do they hate kids if they were one?" Well, she responded with some very, very good answers. Now, you will notice from looking at this, what is happening to her writing. Now, she's really writing. She's talking to me, genuinely talking...My kids that are successful are confident enough and they're secure enough with themselves that they can take the risk of becoming involved with a book. I've got kids who aren't anywhere near that level of security. These are adolescents who are going through an awful lot. I've got kids who are concerned with their peers, that if they read and really made that involvement...and with me, too...they'd consider themselves pansies. I've got kids it's just not macho enough to do this kind of thing. There's a real peer issue...it's very, very difficult for some of the kids.

Different projects are grades as tests...The Book Booster is one...Every time they read a book, they're supposed to make out a "summary card" which is the author, title, and a summary of the book, and who might like it...Then we'll have our silent reading. I will ask them to take out their Reader Response Booklets and to draw conclusions. "What will happen next in your book? Why do you think it will happen, based on facts, clues, experience, etc.? So they're transferring the understanding of the skill into their actual reading. And then, I'll look at their Reader Booklets to see how well they've actually understood the skill.

One teacher explained how the Reader Response activity works, a favored approach in several literature-based classrooms:

> I will put up open ended questions on the overhead. For instance, the planet Uriel [A Wrinkle in Time]. One of the questions might be, "If you were on the planet Uriel, what kind of place do you think it would be?" Then, based on their reading, they had to write that down. All other times, they can just go ahead and write something.

Another teacher described how she arrives at a grade for the marking term for the Reader Response activity of a student's overall grade for Reading, and in so doing, reveals the personal style that each teacher may shelter in their record books:

> When I don't think they've answered, they will get an "R", meaning they have to go back and write a little more. Then it is grades "4,3,2, or 1", "4" being the best and "1" being the poorest. So, at least for those students who don't like to write...at least they're getting a "1" for putting pen to paper. Hopefully, those "1"s will increase into "2"s, and so forth...I probably do anywhere form 12 to 15 of these Reader Response activities in a term...I add up the points and then I have a point system of my own...If a student's total points fall within a range, it becomes a test grade on the Reader Response.

In sum, teachers in basal and literature-based classrooms perceived that they used a variety of formative and summative methods to evaluate the progress their students were making in Reading groups. Teachers reported that they continuously observed and questioned students to appraise their knowledge and progress in learning. Some teachers said they used frequent quizzes and drills to evaluate student learning. Section tests were a common summative evaluation tool in basal classrooms, and arts-integrated writing projects were often transformed into a test grade in literature-based classrooms.[26]

Perhaps due to the prevalence of packaged evaluation materials in basal classrooms, evaluation techniques tended to be perceived by teachers as continuous and objective with standards for performance announced or known to students beforehand.

Perhaps due to the evolving nature of many assignments and activities in literature-based classrooms, evaluation techniques tended to be reported as emergent and teacher-made with standards of performance linked closely to individual students and creative assignments. The evaluation methods teachers described were perceived to be suitable for the particular instructional approaches for each classroom. There seemed, however, to be a more substantive match in literature-based classrooms between the evaluation methods employed and the expressed goals for student learning that teachers expressed. (Please see Appendix G for an analysis of expressed teacher goals for student reading.)

Finally, it should be noted that the role of students in their own evaluation was unclear from data reported in the interviews. While several teachers reported efforts to include students in the evaluation of learning and instruction, the overall data for basal and literature- based classrooms suggest that students were seldom asked or taught to participate in evaluating their own learning progress in Reading.

Performance Criteria for Students in Basal and Literature-Based Classrooms

Teachers were also asked in the interview to identify the performance criteria they used to judge student work for students placed in higher and lower ability groups. Teachers were also asked about the similarities and differences in their evaluation methods and performance criteria for students in different ability groups. These interview questions evoked the following results.

Perhaps because several different subjects other than Reading were taught and because during Reading period various assignments were given to students, teachers preferred to answer or embellish their responses during this part of the interview with specific examples. Moreover, several teachers openly expressed that they had not often thought about the substance requested in these questions. Initial data analysis of

performance criteria reported by teachers revealed different criteria being applied for students in high and low ability groups. Thirteen teachers gave a different set of criteria for students in different ability groups for how they determined whether a student's work was successful or needed improvement.

However, closer examination of the data for these questions revealed categories that emerged from the responses of teachers who were in basal classrooms that differed from data reported by teachers who were in literature-based classrooms. Also, the data differed if there were within-class or between class ability groups. When analyzed along these lines, evidence suggests that performance criteria in basal classrooms tend to be closely connected to levelled materials and the items included on the Report Card. The criteria in basal classrooms tended to be predetermined and absolute.

In contrast, performance criteria in literature-based classrooms tended to be more individualized and relative, depending on a student's reading interests. Quality of reading selections, quality of students' oral and written responses, and individual creativity were heavily weighted criteria in novel-based curricula. Two literature-based teachers, for example, expressed frustration at report card time because they thought that grades did not accurately reflect a student's reading achievement.

In one inner-city school with a high percentage of low achieving children from economically poor homes a distinctive local meaning not anticipated by this researcher emerged in the speech data reported for the three teacher respondents. The lower the student's ability level and social disadvantage as perceived by these teacher, the more the performance criteria seemed to be focused on desirable social behaviors rather than academic performance. For example, teachers said that effort was a key factor; whether

or not the student tried hard, was a good worker, was self-sufficient and independent, raised his hand properly, participated politely, and sat still.

Nine teachers in basal classrooms said that their evaluation methods were similar for students of different ability but that their performance criteria differed for students of different ability. These differences in performance criteria tended to be described quantitatively. Teachers in basal classrooms said they spend more time on oral reading and phonics instruction for students in lower groups than for students in upper groups. They reported that lower groups also proceeded at a slower rate, needed to be given more specific directions, and must have tasks broken down into smaller steps. One teacher reported that lower ability students were not being asked more interpretive, higher-level questions during reading instruction. Yet, higher ability students were asked "many more" interpretive, higher- level questions during reading instruction.

Seven teachers in literature-based classrooms said that their evaluation methods were different and their performance criteria were different for students in high and low ability groups. Further, these teachers tended to respond in qualitative terms when referring to performance criteria for their learners. For example, teachers reported that performance criteria used to arrive at a grade was based on getting to know each student and the kind of work he/she was capable of doing. Different academic behaviors were articulated by teachers as the basis for successful performance across ability groups. For example, regarding the content and mechanics of student responses to essay questions, two teachers expected students in all groups to write complete sentences in paragraph form. However, they also had differing performance criteria for students in various ability groups. The teachers reported these differences in performance criteria:

lower ability group students - provide basic answer, little elaboration

higher ability group students - answer question more specifically, elaborate, use question to make personal references, reason logically

Another teacher was clear about performance criteria. This teacher said that nine criteria set the parameters for classroom performance: 1) more detail; 2) thorough understanding of what was read; 3) proof reading; 4) neatness; 5) content and quality of student answers; 6) clarity of student answers; 7) back-up, support provided for their answers; 8) well-written summaries; 9) accuracy in spelling. The teacher said that for highs, more of the above was expected; for lows, less of the above was expected.

Several teachers recognized that high, middle, and low distinctions were relative ones and may even be overstated for their particular pupils. One teacher using a literature-based curriculum and within-class reading groups responded:

> The three groups are a little more evenly matched to the extent that they are able to elaborate more, bring in more details, make more references, glean more impressions from the book than the less able...My expectations are still the same. I still expect them to write in complete sentences, to answer information about facts or detail correctly, but how much is put into it, the wealth of it, is what would be reflective of the differences of the three groups.

In sum, all teachers interviewed seemed to expect more quantity and more quality in performance from students in higher ability groups than from those in lower ability groups. However, it was not clear that assignments for higher and lower ability students were markedly different in either basal or literature-based reading environments.

The data for this research question suggest that teachers perceive themselves in the somewhat contradictory position of stating high expectations for all their learners but applying different performance standards for students in higher and lower ability groups.

Some descriptive studies on tracking (e.g. Schwartz, 1981; Oakes, 1985; Dar and Resh, 1986) and classroom literacy (Borko and Eisenhart, 1989) have indicated that classroom learning environments are substantively different for high and low achieving students. Interpretations of teacher perceptions for the present research, however, tend to support findings by Carbo (1987) and Good and Brophy (1987) suggesting that there are quantitative distinctions among the tasks assigned to students in different ability groups, but qualitative distinctions in what students of high and low reading ability are engaged in or assigned to do in the classroom are not verbally distinguished by teachers and may not be significant.

Criterion-referenced and comparative-referenced evaluation. Teachers were asked whether they viewed their approach to pupil evaluation as primarily comparative-referenced or criterion-referenced. The view set forth for teachers during the interview was that when a student's test or academic performance in general is interpreted in relation to the performance of others in the defined group, the resulting score or grade is said to be comparative-referenced. Hence, approaches to classroom pupil evaluation that essentially describe how well a student has performed in relation to others can be referred to as comparative- referenced evaluation approaches.

And, when a student's performance is interpreted in relation to defined content, skills, attitudes, or behaviors expected by the teacher, the resulting evaluation is said to be criterion-referenced. Approaches to classroom pupil evaluation designed to describe how well a student has performed in relation to specific content and skills being taught can be referred to as criterion-referenced evaluation approaches. Although unfamiliar with the terms describing these two approaches to evaluation, teachers quickly grasped their meaning.

Teachers were divided between describing their approaches to classroom pupil evaluation as primarily criterion-referenced and a combination of both criterion and comparative-referenced. Nine teachers said they evaluated students on a criterion-referenced basis. These teachers offered statements that reflected a belief in the importance of judging how far a student had progressed since the beginning of the school year rather than judging how well the student stacked up against other students in the present or past classroom group. For example, one teacher stated:

> I like to evaluate them against themselves. Especially, since there is such a range, not only in their reading ability, but motivation. Some of the best readers can be the least motivated...If I don't think a kid is working towards his capabilities, I'll just show it and say "Look at this work". All of their writing is kept in the folder, and they forget that. I asked them the other day to pull out the first piece of writing they did. They were amazed at what awful writers they were in September. So, they see their own progress when these things are kept.

Some of the teachers choosing "criterion-referenced" seemed to be unsure about their overall approach. One such teacher's response shows this confusion:

> I think it's the criterion-referenced. Definitely. I really don't think I make comparisons between students.

Two teachers said they used a comparative-referenced approach as their primary basis for student evaluation. The following comment illustrates the perceptions.

> I guess I have a comparative-referenced approach and I guess it's based upon the curriculum and what I've been teaching over the years. So, as far as reading goes...the kid, if he's on a 6th grade level, he's got to be in the 6th grade reader, and he's got to be handling it...But, I'm not thinking about that consciously when I'm working with all the different groups. Basically, I'm thinking about how they're doing against one another in their groups, or how they're measuring up with one another. Usually, I'll really notice when someone sticks out. So, maybe I am basing my judgment on that.

Sometimes comparative-referenced approaches are broadened to reflect performance assumed to exist elsewhere in the community as in the following teacher's response:

> All of the teachers as you work your way up through the grades, even in the middle school, we all have that type of view of the general group. You have to have some sort of standards within that group. I would say we compare them to the other students in the building, which would be a whole different ballgame if you were at _____ school. Then your standards have changed. So, our acceptable standards, no doubt, have to be lower than in more affluent neighborhoods and where there are more motivated kids.

Eleven teachers reported a "mixed bag" preference or combination of criterion and comparative-referenced approaches to student evaluation. These teachers were not sure about how much emphasis was placed on which approach, but they tended to say that it was weighted in favor of criterion-referenced.

> I think, both. But, it's not a 50-50. I am first confronted with the child from yesterday to today. But, there has to be some comparison with the grade level...60-40 is good. Sometimes even 70. I start with a kid and accept him as he comes in to me as is, and then I work from there. Now, remember, a lot of what we're doing in Reading is reflected in our writing...So, what I do is pick an area and stress it. For example, "point of view". If I'm doing "point of view" activities in Reading then "point of view" will come across in the writing.

> Theoretically, I would like to think that I evaluate the child against himself. Hopefully, I know the child well enough that I know what his abilities are, and is he working up to his capabilities. But there are times when you have to come up to a situation where this is what the group is expected to do, and we're going to measure everyone up against this...So, 70% of the first, 30% of the second. I would like to think, mostly the first, being measure the child against himself. The second being, this is a standard we have to hold to and we're going to measure everybody against this...so and so got 100%, and if you got 60% you didn't do very well.

Some teachers expressed frustration about contradictions between school district grading policies and their own views toward student evaluation. For example, one teacher stated:

In the school system here, we have a grading system but we're supposed to compare all the fifth graders across the school system and then arrive at a grade and, only 20% of the children in your class are supposed to have an A and so on and a lot of us think it's positively the most ridiculous grading system so when it comes to making the comparison on the creative level or the critical thinking level I'm comparing the child with him or her self and the progress that they make during the year. I don't make comparisons with kids in the room...I'm just interested basically in each child's progress and at the beginning of the year, middle, to the end.

This same teacher raised a crucial issue that suggests teachers may desire to create their own assessments of pupil learning, but also realized that it could be a formidable task.

As a Carnegie school, you know, we are allowed, although I think it would be a horrendous undertaking, to stop all statewide testing and standardized testing and design our own test to assess what we are doing in our program here. I'm an advocate for that. That's what I would like to see.

Finally, one teacher perceived that assignments given to students are often explained in criterion-referenced terms. However, when grading, teachers may single out the best papers and use these as a basis to grade all the others.

It depends on the assignment that I give them. For example, I just gave the student's three questions in class. They were all really critical thinking types of questions that I asked them to write their responses to. I will grade those according to how well they have used the details in the book to give me their answers. I am looking for specific points for them to discuss...Then I will compare the students' answers and give them a grade.

In sum, the concepts of "comparative-referenced" and "criterion-referenced" are understood by teachers but are not generally viewed as mutually exclusive. Twenty-one of the teachers perceived their approaches to student evaluation to be either criterion-referenced or a combination of the two with more emphasis placed on criterion-referenced. Although only two teachers described their approaches to student evaluation as comparative-referenced, the remaining teachers may have found it difficult

to reveal their actual evaluation approaches. It is fascinating to hear that 21 of 23 teachers in the present study view their primary approach to student evaluation in relation to specific content being taught and personal growth of the learner, contradicting an assumption that too many experienced teachers employing within-class ability groups, consciously or not, often apply comparative-referenced approach when evaluating student work, grading them for a marking term, or making recommendations for promotion and retention.

Comparative-referenced approaches to evaluation at the classroom level may have significant implications for students that are in conflict with responsibilities concerning equity and sensitivity to individual and cultural differences of children. These are responsibilities which link the role of education with a democratic society. The persistent use of comparative-referenced evaluation may subvert opportunities for students placed in low ability groups to be fairly assessed and for higher ability students to be sufficiently challenged.

In providing an overall summary for the results of Research Question 2, teacher evaluation methods and performance criteria in reading vary depending on whether they are in a basal or non-basal reading curriculum. Teachers reported that they expect students in higher ability groups to move at a faster pace and produce more detailed written work than students placed in lower ability groups. Teachers said their approaches to student evaluation attempted to determine a student's growth in performance in relation to where he or she started at the beginning of the school year rather than comparing a student's performance to fellow classmates. Experienced teachers in the public school reading classrooms of the study also recognize that learning progress in reading is influenced by a host of cognitive and affective conditions.

Teachers in basal classrooms tend to design evaluation systems that provide them with information that is in accord with what must be reported on report cards and that can be used to justify grades to parents. Teachers in literature-based classrooms tend to design evaluation systems for these pragmatic purposes as well, but to also provide information about the quality of reading that the child is engaged in and how well he/she is able to think, read, and learn independently. Both groups of teachers perceive a swell of psycho-social factors as impinging on the child's interest in reading. Some teachers seem overwhelmed by these influences, and some seem to be converting them to advantages in helping the individual child to learn.

Research Question 3:

What are the Ways Teachers Use Information Emerging from their Evaluations of Student Learning in Reading?

The data for answering this research question are organized into three parts. The first part describes various ways teachers say they use data from their evaluations of learning to communicate with parents and students and to improve instruction in Reading. The second part considers the ways teachers report they use evaluation data to make placement decisions of students and the nature of group re-assignments that occur in classrooms where ability grouping for Reading is employed. The third part analyzes how teachers use evaluation data in record-keeping and to derive a Report Card grade for Reading for individual students.

Ways of Using Pupil Evaluation Data to Communicate and Improve Instruction

Interview data of the 23 teachers in this study indicate that their use of data from pupil evaluation in Reading to communicate and improve instruction can be described in seven categories. These categories are: 1) To prepare written reports of

student progress; 2) To communicate with parents; 3) To inform instructional decisions; 4) To provide feedback to students; 5) To learn more about the needs and interests of students; 6) To facilitate classroom management procedures; 7) To evaluate instruction. Teacher responses in these categories were not evenly distributed. Given the limited number of teachers participating in the study and given the interpretive nature of the study, it is not appropriate methodologically to make quantitative summary statements. In the discussion and analysis for this chapter, numbers have been presented as descriptive of the corpus of data rather than to suggest conclusions. The data analysis for each of the seven categories follows:

To prepare written reports of student progress. Fifteen of the teachers interviewed responded that they used evaluation data on students to assign or justify grades for report cards, progress reports, and other such summaries of student classroom achievement. Three formative and two summative strategies help to understand how teachers used pupil evaluation data for Reading in this category. First, the majority of teachers cited preparing grades, marks, and comments for report cards as their primary task stemming from classroom evaluation. They used data from student evaluation to guide them in this significant task. Report cards were generally issued four times per year or about once every eight to ten weeks. Similarly, in four schools in the sample, teachers had to prepare Progress Reports on their students. These were produced in the middle of each marking period in intervals that corresponded to the dissemination of report cards. These reports were more narrative and abbreviated in format than report cards, which usually followed in four to five weeks.

Teachers also mentioned reports of student progress that they maintained in formative ways. For example, daily sheets or daily "report cards" were issued for

students who were having behavior problems in the classroom or for whom special learning targets had been established. These sheets were sent home daily to parents through the student. These reports continued until either the parent or the teacher or both decided they were no longer necessary.

Three teachers kept a journal of the progress each child was making in Reading. These journals contained anecdotal information about what was happening with the child at school and sometimes at home. Home information believed to be pertinent by the teacher might include special medication, sleep and study habits, and responsibilities for the care of other siblings while parents worked. School information deemed important to document included peer group interactions, lesson attentiveness, assignment completion, perceived personal qualities, and the like. Teachers reported that the actual kinds of information recorded in journal entries might vary from student to student.

Finally, three teachers said that as data from student evaluation were gathered, they were transformed into a mark or grade and merely recorded in gradebooks. Marks recorded in gradebooks were then used to form the basis of a student's grades for a report period.

To communicate with parents. In addition to report cards and progress reports, seventeen teachers said they used evaluative data to get ready for parent conferences. These conferences typically occurred once a year in October or November. There were other opportunities taken by teachers, however, to converse with parents about how their child or the class as a whole was progressing. These included phone calls home, written notes, class newsletters, and parent visits to the school.

One teacher said that during the first several weeks of school, calls were made to the parents of all the children in the class to learn as much as possible about each

child before intensified instruction was begun. Other teachers reported that they made phone calls and sent written notes to parents, usually when there was a problem occurring at school. Two teachers had instituted a class newsletter that the teacher and students developed together to serve as a communication tool about class activities and upcoming events. One teacher periodically sent home a teacher-authored newsletter for this same purpose. Teachers said that newsletter formats were often used to invite parent participation in a special event, such as a play or field trip that the class was planning.

Four teachers said that they arranged for parents of some students to come in at times other than parent conference night. They reported that the parent conferences scheduled by the school were too infrequent to discuss the exigencies relating to learning problems of particular students. And, they said that when, on occasion, a parent of a student volunteered at the school or in the classroom, the visit presented an opportunity to talk about the child's progress.

Finally, data for this category revealed frequent complaints among teachers about the lack of parent involvement in the education of the child. These complaints were voiced most in urban schools serving poor children. Teachers did not mention arranging personal visits to the homes of children to discuss their learning with parents.

To inform instructional decisions. Thirteen teachers reported they used evaluative data for this purpose. These teachers identified ten ways that they use data from pupil evaluation in reading to inform subsequent decisions they might make regarding their reading curriculum. Five uses are for formative decision-making.

First, when evaluation data showed that an individual, group, or the whole class was having difficulty or had failed in their learning, teachers often determined whether

to review, remediate, or reteach the content. Eight teachers mentioned the use of flexible, ad hoc groups to accomplish this purpose. Second, formative evaluation data were used to change the directions for an assignment, to change or omit a lesson, or redesign a segment or episode of instruction altogether. Third, whether to move on to the next content objective or skill was a decision with which teachers in basal classrooms were particularly concerned. They perceived that their evaluation strategies provided them with information upon which to make this decision. Fourth, 15 teachers said they used data to continue to determine ways to better assist student's learning. Five teachers said they individualized materials and assignments in Reading to address this purpose. Two teachers admitted that pupil evaluation data revealed that additional help was needed to better address some of their students' learning needs and sought help from the Reading Specialist in this regard. Fifth, four teachers said that they used evaluation data to guide them in getting students prepared for the next grade level. This was a high priority for some teachers as they decried the lack of preparation of many of the students they received from previous grade levels. These teachers reported they felt both pride and pressure to ensure that their pupils went to the next grade ready to meet its demands.

Data from evaluation of students were perceived to be useful for five summative decision-making purposes. First, teachers made minor revisions to larger segments of the reading curriculum, such as altering the sequence of a particular textbook unit for a reading group in a basal classroom or changing several of the whole class reading selections in a literature-based classroom. Second, teachers made temporal decisions of when to move on to the next unit of content and qualitative decisions about how much emphasis to give to the various topics within the unit. Third, teachers reported that

they sometimes set individually tailored goals for students when their evaluations revealed that this was an appropriate course of action to correct learning problems. Such tailored goals might be formulated for lower achieving students having difficulty and higher achieving students requiring more challenging and faster-paced objectives. In conjunction with individual goal-setting for students, some teachers pressed their energies further to change or modify learning materials for these students as well. Fourth, three teachers said that as a result of evaluation data they might change the directions for an assessment to be given to students. For example, one teacher said students were permitted to take open-book quizzes and tests, because this approach would make text-anxious students more relaxed and thereby enable them to provide answers that were nearer to their true knowledge of the material. Finally, some teachers used pupil evaluation data to make changes in their own Reading curriculum for the following year such as vanquishing certain assignments, books, projects, activities, and learning experiences altogether and replacing these with new ones. Four basal classroom teachers recommended to principals that changes in textbooks be made and that between-class ability groups be re-organized so that teachers can use evaluation data to group students within their classrooms.

To provide feedback to students. Teachers report utilizing evaluation data for three formative and three summative decision-making purposes to provide feedback to students regarding their learning progress in reading. First, teachers said they communicated regularly to students about how well they were performing. Communication of this kind was usually verbal. Second, teachers said they wrote instructive comments on papers and homework assignments. This form of feedback included comments on class diaries and journals, and a variety of other independent

writing assignments. Third, three teachers reported they used evaluation data in deciding to hold private conferences with individual students to discuss learning progress.

Four teachers mentioned using evaluation data for summative purposes of exhibiting student work, with an emphasis on displaying exemplary products, projects, papers, essays and poems. Second, three teachers reported communicating test results to students in the form of grades or numerical scores. These tests most often were standardized tests of the comparative-referenced achievement variety or criterion-referenced variety as produced by basal publishing companies. Teachers rarely mentioned developing their own summative assessments to evaluate student learning. An exception was a teacher in a literature-based classroom using an interdisciplinary thematic approach to reading instruction. This teacher developed tests to assess student learning at the end of literature-history units. The tests were designed to measure student knowledge and attitude formation stemming from the books they read. Third, teachers in literature-based classrooms especially, reported giving culminating projects (e.g. mythology writing, creative book summaries, script writing for a TV mini-series). Three of these teachers reported that feedback on these various projects was provided in writing or orally during individual conferences with students.

To learn more about individual students. Fourteen teachers reported ways that they use formative evaluation data to develop personal rapport with their students and to better attune themselves to their students' needs and interests. Five strategies were indicated. First, teachers used data to internalize judgments about a student's ability and classroom accomplishments. Second, teachers used data to monitor intrinsic and continuing motivation levels of individual children. Third, teachers reported using data to help students having difficulty feel less overwhelmed and more confident in their

learning. Fourth, teachers responding in this category said their evaluations of students helped them discover special interests and hidden talents of children. Information gleaned from this context often made teachers feel full of new possibility for their students and nudged them forward to try to surface and build upon these talents and interests advantageously during instruction. Fifth, as mentioned before, teachers kept journals and arranged individual talks with students.

To facilitate classroom management procedures. Only three teachers made responses relative to this category. Nonetheless, their comments are instructive and may provide directions for further inquiry and practice. For example, one teacher used evaluation data to monitor and orchestrate social relationships in the classroom. Another teacher said that evaluation data were used to identify and resolve problems among students. Finally, two of the three teachers indicated that they used formative and summative evaluation data on the academic and social behaviors of students as a basis for distributing rewards and punishments to individuals, groups, or the class as a whole. They reported that such rewards may take the form of verbal praise, reduction of work assignments, and spotlighting of student work. Punishments may include verbal admonishment, loss of in-class privileges, notes home to parents, and after school detentions.

To evaluate instruction. Four teachers specifically reported using data from student evaluation to evaluate their own teaching. First, teachers said they talked with other teachers to determine how their "kids seemed to be doing in relation to kids in other classrooms." Sharing with other teachers enabled them to improve strategies for dealing with individual students, and to get ideas for "adding more spice" to their curriculum. Second, teachers provided forums for students to share their ideas for

improving instruction. Third, two teachers said they used the Metropolitan Achievement Test (MAT) as an indicator of student growth for the year and a potential barometer of successful instruction.

Following are a sample of teacher comments to illustrate the way they use evaluation data to communicate with parents and students and to improve instruction:

> Obviously to evaluate students in terms of the report cards, give feedback to students in terms of progress, feedback for me for redesigning planning lessons, curriculum, when do I go back and review something, when do I need to skip something, placement in reading groups, etc...I guess also looking at things emotionally for the kids, just getting a feel for the student, as a barometer for the student's emotions finding out about interests, their own interests beyond the classroom walls that may not get addressed by the interaction between you and the student, records for parent/teacher conferences. And that was the last one I want to deal with, using the information in dealing with peers and the social relationships.

> I might group small groups together and go over things with students who are having a problem. The other part of the test that I didn't bring was the skills, and it's sometimes quite obvious that the student didn't get the skill from evaluation. I will go back and go over it with them, give them additional work on it, and so forth...Or if I find that the whole class had a problem with one particular thing I might have to go back and go over that again.

> Obviously, for grades. I also use it to help get them read for the 6th grade...With my 5th graders for spelling and math, I do carry it over to my other subject areas because I like to correlate all the subjects. It's hard for them to accept that you need to have really good reading skills to do math. The reading skills continuously overlap with out spelling skills.

> Well, for the report card...And certainly parent conferences. Parents are very grade conscious and they want to know exactly...not is he doing better or worse...but what is his grade. And it goes on their records.

> I record it. I keep a journal and I see the journal as part of the evaluation process. The boys and girls have their own reading journal too...My journal will tell me if a student is, for example, going through a difficult time at home, if there's a separation or death in the family, if there's a sickness in the family, if the students are absent from school because of a vacation or illness. Some particular students seem to be

enjoying a book by a particular author more than other students are enjoying it. It's kind of like my own personal diary.

We have a display case in the center of the main lobby. I use that for exhibits of their reading projects. Parents come in to view reading projects that we do.

If I feel the children are deficient in some area, I use my evaluation methods to discover that and then I will re-think my own teaching of that skill, reinforce it with the child individually.

Well, of course, it will be one criteria I use if the child is going to repeat the entire grade, which I don't like to have happen in grade 5...I feel that if they're going to repeat a grade, it should have happened a long time ago. But that is one criteria I will use for promotion...their reading level or ability, or their progress through the reading level...I particularly don't like to move children down because I think that's devastating to their ego within a classroom. They're very aware of which reading group is at what level even though you try to call them the Blue Group, the Red Group, or whatever. I never try to move a child down. If they're having skill problems, I'll try to do it at the end of the year and then it will be a whole new group of kids and nobody will know whether they were in that book last year or not. I also try to give them as much individualized help as I can so that they're holding their own at whatever level...unless they've been out sick or something like that. I'll go to the tutorial program or the city reading person to help keep their head above water, if you will. I try very hard never to move a child down in this class. It's devastating to a child to be moved down.

Basically, their day to day work lets me know whether they're going to keep going on with the next section of the book. When we test on one part of the book and we test on a magazine, again, how they do on the test determines whether or not we'll go on to the next part of the book. It's pretty cut and dry. Again, if the group as a whole didn't do well, there will be a lot of review on just the skill. Very seldom have I had to go back and review a whole magazine. But, a lot of the time, I will have to review some of the skills. If I see, for example, that seven out of the eight kids in the group didn't pass this one particular skill, then I would review...Sometimes, it's used to move people to other groups or move them down.

If I get real good work from one or two students, I will use it as models for the rest of them...If I give a student a "B" on a Book Booster, then I would mark it down as an 85%...Sometimes I will have them do it again...By showing off their work all over the room, is another way.

I use it to evaluate their growth in that particular area. When you've had a swiss cheese approach, and the kid's reading is very frayed, you

look back at those holes. My whole purpose is to give them strategies to fill in some of those holes.

...using the amount of independent reading that the student has done outside...they have to keep a record of their independent reading with me, and they have to do some very simple work with me to show me that they have read the book. I read their reading journals and I use the amount of outside reading that they've done. I use the reading test and the reading project at the end of the book that they've done. If they've had a lot of difficulty in their own life during the quarter but their effort has still been there, then I add a couple of points to their numerical grade.

In sum, teachers report that they engage in a range of evaluative tasks and use evaluation data in a variety of ways. These student evaluation activities would probably be difficult to observe directly since they are integral rather than part and parcel to other teaching functions. Moreover, evaluation behaviors of teachers occur continuously. Research suggests that teachers make, on average, an interactive decision every two minutes (Clark and Peterson, 1986). Evaluation data emerging from these. evaluation tasks lead to further re-cycling of data for formative and summative decision-making purposes.[27] Twenty-two of 23 teachers said they use evaluation data primarily for making out report cards and/or preparing for parent conferences. However, when probed further, a more complete picture emerges of the ways teachers use results of student evaluation. Seven categories of how teachers use pupil evaluation data in Reading were described.

Ways of Using Pupil Evaluation Data to Make Placement Decisions and to Move Students Across Ability Groups for Reading

Teachers reported using evaluative information to make placement decisions of students. First, all 23 teachers reported they used numerous and varied criteria during at least three key interval periods during the academic year to form instructional groups

to teach Reading. Second, 18 teachers in the study said they move from two to eight students a year from one ability group to another.

Further, teachers reported they refer a student for special instructional services such as those provided by Chapter I programs, resource room teachers, and one-to-one tutoring; use student evaluation data to make recommendations to next year's teachers in establishing initial reading groups; and, make recommendations to promote or retain a student.

Prior assumptions about the positive effects of ability grouping for lower-tracked students are being seriously questioned by researchers and practitioners (Allington, 1983; Slavin, 1987; Massachusetts Board of Education, 1990). Increasing numbers of students and parents are challenging the need for ability groups and refusing to accept sorting as a necessary condition for learning. One of the ways teachers reported that they used data from evaluation was to alter grouping assignments in Reading.

There were 515 students represented in the classrooms of teachers in the study. Of these, 96 or approximately 1/5 were moved from one group to another during the academic year. The range in number of students moved from one ability group to another was 0 to 12. Two teachers said that no students had been moved, one teacher said that one student had been moved. Conversely, one teacher reported movement for 10 students and another teacher reported movement for 12 students. See Table 7 for a summary of data describing teacher reports of the number of students moved from one ability group to another by all teachers interviewed.

The teacher responses were further organized along two dimensions of whether the movement was to a higher group or to a lower ability group. Results indicate that 15 students were moved up and five students were moved down in within-class grouping.

Table 7

Bar Graph Depicting Number of Students Moved in Reading as a
Proportion of the Class Total for the 23 Classrooms in the Study

No. of pupils in the Classroom

Reading Classrooms in the Study Sample (N-23)

Number of students moved to different
reading level group
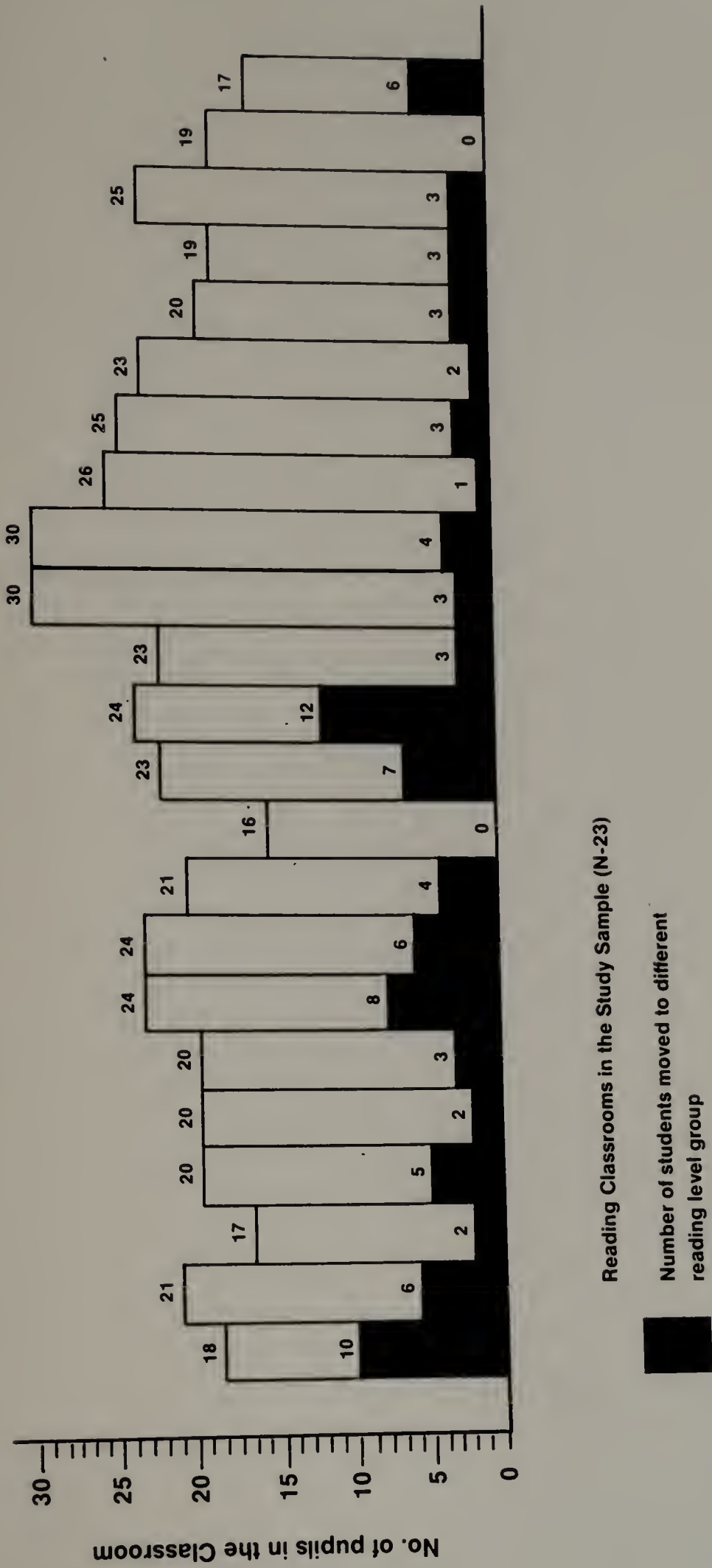
Between-class movement, however, indicated that five students were moved to a lower group and four were moved to a higher group. Finally, when students were referred forspecial education, Chapter I, or Resource Room services, the movement in ability group was classified as downward. Data generally indicate that when students left the class they mainly went to receive remedial instruction in Reading from a Chapter I or Resource Room teacher (N=17). Only one teacher reported that one student in this classification was moved "up" after being sent out of the classroom for special services and this student was returning to his/her original classroom. Table 8 summarizes these data reported by teachers.

In sum, within-class movement data is weighted in favor of upward movement to a higher ability group. Between-class movement data is varied, with a fairly even distribution of movement to a higher and lower ability group. Out of class movement data is intrinsically downward, with most children going to Chapter I or the school Resource Room for special help in Reading. Overall, however, there is little student movement across ability groups for the reading classrooms represented in the study.

Using Evaluation Data to Determine Student Grades for Reading

All teachers interviewed reported that evaluation data were used to determine student grades for Reading. Teachers willing to share how they used evaluation data for grading purposes spoke freely and openly. Interview data and analysis of teacher gradebook sheets for Reading uncovered six significant themes or topics in the domain of use of evaluation data for the purpose of recording, marking, and grading student performance. These topics include: 1) composition and format of teacher gradebook sheets for Reading; 2) factors that influence grading; 3) methods for recording data; 4) methods for arriving at a term grade; 5) role of intuitive, mentally stored data; 6)

Table 8

Number of Students Moved to a Higher (Up) or Lower (Down)
Ability Group in Reading Classrooms within the Study

156
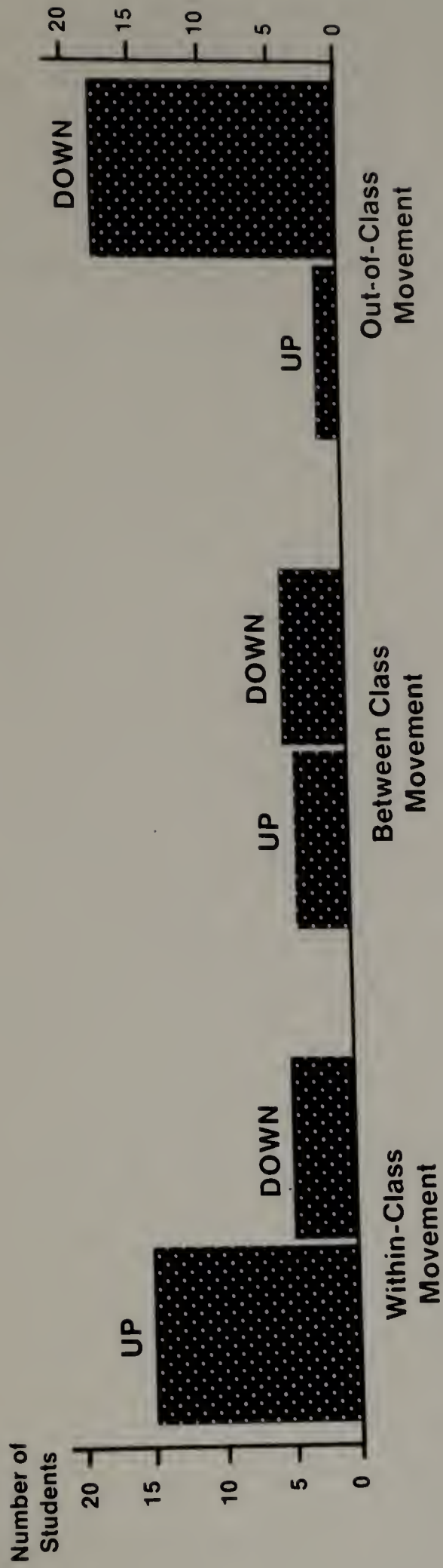
academic improvement and grading. Selected information from teacher interviews that is pertinent to these topics is described next.

Composition and format of teacher gradebook sheets for Reading. Sixteen teachers either shared or described their Reading record book sheets for the interviewer. Those record sheets reviewed directly were found to be neat, tidy and well organized. Students names were typically listed alphabetically down the left-hand margin and organized by high, medium and low ability groups. In columns across the top of the sheet, were the indices or categories used by the teacher to measure performance during the marking period and to determine a final term grade for each student. These categories varied widely for teachers in the study, each having several that differed from the next colleague.

In basal classrooms, a composite of common categories emerged from hard copies and verbal descriptions of teacher gradebooks. These included selected "skills" columns, comprehension columns, selected test and quiz columns, a column (or separate chart) to monitor oral and independent reading progress, and a column to keep track of attendance. For consistency, most teachers said they arranged their gradebooks to match the Report Card format and its requirements. It was observed, however, that Report Cards came in all different shapes, colors, and sizes. Moreover, there were a variety of marks, scores, grading criteria, and performance level criteria indicated among them. Finally, teachers said they recorded anecdotal information about individual students. These notes ranged from brief comments summarizing a student's progress for the term to more extensive narratives maintained in journals, entered as often as possible by some teachers. Table 9 shows a composite gradebook sheet for a teacher in a basal Reading classroom.

For teachers in literature-based classrooms, the gradebook sheets were organized in similarly tidy formats. In addition to a column to check oral and independent reading progress, the column categories tended to include grades for special culminating projects, journal grades, test results from a particular novel read, and other assignments preferred by the teacher such as reader responses, book boosters, work on computer programs designed to reinforce reading and writing skills, etc. Teachers' gradebooks did not often reflect a column to record growth in comprehension, skill, and vocabulary development as did basal-based gradebooks. Table 10 shows a composite gradebook sheet for a teacher in a literature-based Reading classroom.

Expressed Factors that influence grading. Teachers were clear and particular about the elements of student performance that influenced their grading practices. These elements included academic and non-academic factors. Interestingly, teachers were more agreed in their non-academic sources of grading than their academic ones.

Academic sources of grading included achievement indicators in a variety of reading and writing skills. However, teachers stressed different components in these broad skill areas. Some teachers stressed mechanics and content; some emphasized analytical thinking, work facility, and writing skills. Some teachers stressed worksheets related to comprehension, discrete skill, and vocabulary development, and some teachers were heavily influenced by the quality of reading selections of a student and the motivation of a student to read at other than assigned times. Teachers in basal classrooms gave extra weight to results on section tests and teachers in literature-based classrooms weighted their project assignments more heavily. No two teachers weighted any of these academic elements in the same manner in determining the final term grade for students. Moreover, as described in prior analysis for Research Question 2, teachers

Table 9

A Composite Gradebook Sheet for a Teacher
in a Basal Reading Classroom

	Names*	Comments	Seatwork Assign. (Skills)	Workbook Pages (Skills)	Homework Assign. (Skills)	Unit Tests	Quizzes	Oral & Independent Reading Progress	Attendance	Term Grade
"Tapestry" (High Group)	1. Terry									
	2. Karen									
	3. John									
	4. Kathy									
	5. Martha									
	6. Tim									
	7. Marion									
"Cloud Seekers" Middle Group)	1. Jackie									
	2. Doug									
	3. Stacey									
	4. Jennifer									
	5. Kyle									
	6. Rebecca									
	7. Walter									
	8. Jill									
"Sunshine" (Low Group)	1. Louis									
	2. Chad									
	3. Amy									
	4. Colleen									
	5. Kelly									
	6. Josh									
	7. Robert									

*Student names are fictitious

Table 10

A Composite Gradebook Sheet for a Teacher in a Literature-Based Reading Classroom

Names*	Oral & Independent Reading Progress	Journal Grades	Project Grades	Essay Test on Novel e.g. "Push Cart War"	Skills Assignments (Reading and Writing)	Book Boosters	Attendance	Term Grade	Comments
"Tapestry" (High Group)									
1. Terry									
2. Karen									
3. John									
4. Kathy									
5. Martha									
6. Tim									
7. Marion									
"Cloud Seekers" Middle Group)									
1. Jackie									
2. Doug									
3. Stacey									
4. Jennifer									
5. Kyle									
6. Rebecca									
7. Walter									
8. Jill									
"Sunshine" (Low Group)									
1. Louis									
2. Chad									
3. Amy									
4. Colleen									
5. Kelly									
6. Josh									
7. Robert									

*Student names are fictitious

expected more quantity and quality in these aspects of reading performance from students of higher ability than from students of lower ability.

After a teacher's academic or achievement indicators in reading were tallied and averaged, she/he then factored into the grade several non-academic or extraneous variables such as class participation, effort, attitude, attendance, and social behavior in and outside the classroom.

Almost every teacher interviewed said that a student's effort, attitude and participation in class were taken into consideration for the final term grade. Several teachers mentioned how these non-academic variables influenced their grading. One teacher said "extra-credit" was given for class participation. One teacher claimed, "I'm a hard grader if a kid is lazy. I am a sap if the kid is working hard, trying." In a similar way, another teacher said, "I am a hard marker, but in borderline cases I give the benefit of the doubt to kids." The latter two teachers were teaching poor children in inner-city schools. Interestingly, these same teachers complained about inconsistent and lax grading policies of other colleagues.

Methods for recording data. Teachers grading practices may be further "contaminated" by inconsistencies in recording student performance data in grade books. What may appear as a jumble of hieroglyphics to the outside observer is perceived to have quite clear meaning to the teacher-evaluator. Teacher record grades for similarly weighted assessments in alpha-numeric forms. A student may receive letter grades for some grading indices and numerical scores for others. Teachers also use a creative mixture of √, √+, √-, 0, x , 1, 2, 3, 4, *, + , - , /, and other notations to indicate a student's mark on a particular assignment. Four teachers also mentioned that the scores in a gradebook are selective and not inclusive of all the work done by students over the

term. To illustrate, Figure 2 provides a sample excerpt for a student from a teacher grade sheet.

Names	Sel. Skills Assignments	Homework Assignments	Workbook Grade	Unit Tests	Class Partic.	Term Grade
1. Joe*	✓, ✓+, ✓-	✓+, 0	B+	84, 89	+	B+

Figure 2

Excerpt from a Teacher's Gradebook Sheet

*Student's name is fictitious.

Methods for arriving at final grade. Teachers' methods vary in the way that they combine information from their recorded data and (frequently non-recorded) data such as effort, attitude, class participation, etc. to arrive at a final grade for students. Eleven teachers said that they "average" all the marks, scores and grades, then transform them into a single letter grade for the term. Two said they "eyeball and average," almost in an instant, the various marks given for skills exercises, homework assignments, and seatwork papers. Another teacher "looks across and averages" the grades for the final term grade. Tests, projects, number of books read, journals, and the like were found to be weighted similarly and given a letter grade. The weight of these items in accounting for the final grade, however, varied across teachers. Thirteen teachers divided their gradebooks into 3, 4, or 5 categories, equally weighted, averaged the scores or grades from these, factored in behavior, effort, and attitude, and determined a final grade for the student.

Role of intuitive judgments in grading. Teachers' stored mental data and intuitive judgments of student progress may play a more influential role than all others in grade determination. Two teachers openly alluded to this possibility. "An awful lot

of information about individual kids is stored in my mind. I use this information constantly in working with kids and their parents." Another teacher said, "I keep a lot of stuff in my head. After a couple of months, from what is done in class, a pattern emerges for each child." The teacher went on to explain that students were evaluated on a monthly basis and that intuitive judgments figure often in the evaluation process. Research by Barnes (1985), Thiessen and Moorhead (1985), and others corroborate the prominent role of intuitive feelings in the classroom evaluation of students. Because intuitive skills vary markedly from person to person, more needs to be investigated and learned about this significant dimension of pupil evaluation.

Academic improvement and grading. The many confounding factors shown in these data to be associated with teacher grading practices of students make it difficult to validly interpret a student's grade as an indication of academic improvement, let alone achievement.

In fact, teachers do not often state that grades are intended for this purpose. Most see it as a required task, a task for Report Card time and for justifying student progress to parents. Only one teacher clearly articulated that improvement in academic performance was an important factor in the grading process. Many teachers clearly stated that they were continuously observing whether a student had improved or not during the term. The question could be legitimately asked, however, "Improvement compared to what and for whom?," since goals, expectations, assignments and performance criteria often differed for students in different ability groups. Few teachers reported incorporating students' assessment of their own learning progress into grading practices. Thus, the role of evaluation as a means to improvement remains a "fuzzy" and tangled concept at the level of classroom instruction.

In sum, despite being engaged continuously in numerous student evaluation contexts, teachers make sincere and even gallant efforts to reduce these complex assessment situations into a common cultural symbol of achievement for each student (the final letter grade). Each teacher has his/her own particular methods for transforming concrete and intuitive ratings of a student's academic and social development into grades. Further, although teachers are in the nearest position to determine whether and how well students are improving their reading and writing skills, teachers' grades of students may lack the meaning intended for this purpose. Teachers are not agreed about clear and consistent domains of content that denote reading success. Moreover, some teachers reported that they incorporate a host of variables unrelated to student's demonstrated achievement to assign a grade for that student. These non-academic variables may lead to the formation of "halo" or "noose" effects for particular students and thereby further the distance between a student's perceived and actual achievement. Nonetheless, most teachers said they feel extremely confident that they can justify their grades to parents.

Conditions that Help and Conditions that Hinder Student Evaluation

Prior interview studies have shown that teachers believe grading and evaluating classroom pupils to be the most difficult aspects of teaching (Griffin, et al. 1983). For almost ten years, Richard Stiggins and his colleagues at the Northwest Regional Educational Laboratory in Portland, Oregon have been documenting the areas of neglect and the need for improvement in classroom teacher evaluation of learners. Findings suggest that teachers need and want help in this vital area of teaching. Until more promising assistance arrives, however, it is useful to find out what teachers say are the

conditions that presently help or hinder their ability to evaluate classroom learning effectively.

Helpful conditions. Analysis of interview data indicated that there were three main factors that teachers said supported the student evaluation process for them: 1) Experience; 2) Supportive colleagues; 3) A supportive principal. Each of these factors is briefly described through the testimony given by teachers.

Eleven teachers felt strongly that their experience in teaching was the major factor enabling them to handle the complex demands of pupil evaluation. They believed that their experience enabled them to select and design materials over the years to better assess a child's progress in Reading. Several teachers had "stockpiled" supplementary learning material such as stories, books, kits, levelled reading programs, individual reading inventories (IRIs), etc. Teachers believed that their overall experience allowed them to evaluate quickly yet efficiently. One teacher commented about the relationship of evaluation and experience: "I don't have to think. I just do it." Another teacher felt that eleven years experience teaching gifted children helped in stretching the potential of children and nudging them to set higher expectations for themselves as well. One teacher said that evaluation was made easier as a result of experience: "I do a lot more evaluation in my head. Students tend to fall into categories earlier." In short, teachers said they were able to place more confidence in their judgments about student learning progress due to their experience. In contrast, they suggested that a novice teacher would be overwhelmed in this area of responsibility.

Seven teachers identified the support of colleagues as a second helpful condition for evaluating student learning. The ideas and suggestions, the sharing of materials, the family-like collegiality, the cooperation and encouragement of other staff were all

identified as important supports. Also, team planning meetings as well as informal opportunities to discuss the progress of individual children were viewed as valuable conditions for more confident pupil evaluation.

Six teachers said that a supportive principal helped the evaluation process. Administration was perceived to be supportive when it gave teachers the freedom to teach how they wanted to; when it allowed teachers to experiment and try new methods; and, when it provided leadership and support for professional development.

Hindering conditions. Five hindering conditions recurred in analysis of the interview data for this follow-up question. These were: 1) Time constraints; 2) Report card formats; 3) Grading tasks; 4) Inadequate materials; and, 5) Inherent difficulty in evaluation. Each of these perceived obstacles to effective student evaluation in Reading is briefly described.

Ten teachers in the study lamented the lack of time to gather the full range of information necessary to adequately evaluate students. Teachers said that there was not enough time for Reading itself, the time allocated competed with a full day of other subjects and activity for students. Teachers said there was not enough time for parent conferences. Once a year meetings with parents for a block of 10-15 minutes was believed to be absurd. There was not enough time to individualize materials and instruction, even though teachers recognized that many children in the classroom would benefit. There was not enough time to keep up with the task of grading and returning the mounds of paperwork assigned to students. Record-keeping time was perceived as conflicting with teaching time. There were not enough hours in a day to do both properly. The writing component of the Reading curriculum was viewed as critical, but teachers felt that there was not enough time to evaluate student writing adequately.

And finally, while teachers described a wide variety of formative evaluation methods used to evaluate both student learning and the instructional enterprise as a whole, some felt that there was not enough time to improve instruction significantly *during* the year due to work press.

Seven teachers believed the school or school district's Report Card format forced them into narrow descriptions of what a child had or had not accomplished in his/her learning. In some districts, the Report Card format was not compatible with the structure of ability grouping. Was the child reading at, above, or below grade level? The format did not provide for distinctions in levelled performance. Some teachers also thought it inappropriate to divide a student's language achievement into separate categories on the Report Card. Some Report Cards, for example, required that students receive a separate grade for Reading, Writing, and Spelling.

Seven teachers believed that inconsistent grading practices as well as their own aversion to the task significantly hampered meaningful evaluation of students. One teacher voiced a common sentiment: "I don't like having to give grades. I especially hate to give poor grades." Many teachers observed huge discrepancies in how student performance was marked from year to year, from school to school, and from teacher to teacher. In short, the lack of clear and common evaluation standards supported by a unified system of grading meant that Report Card grades of students were often invalid as true indicators of student achievement and performance.

Six teachers said that they often lacked proper assessment tools or that those assessments that accompanied published materials were inadequate or inappropriate for their pupils. For example, teachers in basal classrooms decried the mismatch between items on section tests and the material assessed in worksheets and workbooks that went

along with the same basal system. Similarly, some teachers felt like they were forcing square pegs into round holes when prescribed materials were clearly inadequate for addressing the range of special needs that children brought to the reading classroom. Some teachers wanted to simply dispose of particular basal series their school district had invested in. In literature-based classrooms, teachers continually searched for quality books for their in-class libraries, but felt that support for these acquisitions was poor, as several said they accumulated books on their own.

Three teachers openly acknowledged that the inherent difficulty of evaluation was itself a hindrance to good evaluation. Two teachers using literature-based reading curricula said that the approach made it difficult to evaluate the progress that individual students were making, particularly in skill development. Another teacher simply stated that, "evaluation is very tough to do, especially in Reading."

Finally, nine teachers made individual comments for this subquestion, that when considered together, provide additional evidence of conditions that hinder effective student evaluation in Reading. These comments are paraphrased from teachers own words. First, when children leave the room for extra help, they miss a lot of class time. This is a hindrance to the regular classroom teacher in maintaining continuity of evaluation of these children. Second, when students are misplaced in an ability group, it jumbles the evaluation process for teachers. Third, when parents are unable to respond to teacher requests for problem-solving meetings, an important link in the child's education is impeded.

Fourth, managing two or more groups simultaneously is not an easy task and can drain the energies of teachers into managerial functions and away from evaluative ones. Fifth, certain ability group structures are viewed as incompatible with certain

instructional approaches - e.g. homogeneous (between-class) ability grouping for a novel-based curriculum. Sixth, children in mid-ability groups often slip through the crevices. Seventh, there often is no additional help, in the person of a teacher's aide or a classroom volunteer to assist in evaluation tasks. Eighth, experience and longevity in the profession sometimes work against themselves creating "psychological ruts" which hinder motivation to do student evaluation well. And, ninth, negative social and emotional factors affecting children's performance in Reading make it difficult to assess a child's real potential.

In sum, teachers identified five major and nine discrete hindrances to effective pupil evaluation in Reading. Conversely, they perceived three factors as facilitating their responsibility in this part of teaching. The demands placed on teachers to help all children learn to their potential are enormous. It would appear that teachers need more resource and professional support if they are going to feel more comfortable in their role as teacher-evaluators.

Chapter Summary

This chapter described analysis of patterns in the interview data for the three major research questions that guided the study. First, it presented teachers' reports of the criteria used to group children for Reading instruction. Second, the chapter described the means teachers use to evaluate student learning and teacher perceptions of the similarities and differences of these evaluation methods for students in different ability-based reading groups. Third, the ways that teachers use data from pupil evaluation in Reading were organized and described. Teachers had a lot to say about these questions and related issues associated with evaluating pupil learning progress in within-class and between-class ability groups for Reading.

In summary, teachers feel that the criteria used to form initial ability groups are beyond their control. They recognized their lack of decision-making for the groups that come to them at the start of the school year yet acknowledged their responsibility as key recommenders of ability groups for the next grade level. As the school year progressed, teachers felt more efficacy about restructuring their own within-class ability groups and used a variety of criteria to do so.

Teachers differed in their formative and summative evaluation methods and in their performance criteria for students depending on whether they use a basal or literature-based approach to instruction. Teachers in basal classrooms perceived their evaluation methods to be similar and their performance criteria different for students in different ability groups. Teachers in literature-based classrooms perceived their evaluation methods and their performance criteria to be different for students in different ability groups.

Teachers reported using student evaluation data primarily to prepare Report Cards and to meet with parents. Teachers also reported using student evaluation data to inform a variety of decisions to improve learning and instruction. Further, teachers used evaluation data to change a student's status in a reading group. Data provided by teachers indicate that student movement across ability groups was infrequent. When movement did occur, teachers usually moved a student to a higher group in within-class structures, either way in between-class arrangements, and to Chapter I or the school resource room when referring a child for out-of-class services. Analysis of teacher gradebooks and record-keeping procedures showed non-uniform but creative ways of grading pupil learning progress for a marking period.

Finally, one compelling interpretation emerging from the corpus of data for the study as a whole is that teachers clearly engage in numerous kinds of assessment evaluations of students' reading. There is great variation in the evaluation practices of teachers. Evaluation data that is collected by teachers does not, however, prompt them to consider alternative grouping arrangements for their pupils nor challenge the prevailing structure of ability grouping.

Data analyzed in this chapter provide a foundation from which to infer important implications about teacher practices in pupil evaluation in Reading. Data analyzed for the major research questions and subquestions also provide a basis for suggesting directions for improved practice and further inquiry in classroom grouping and evaluation.

CHAPTER V

SUMMARY, IMPLICATIONS, AND RECOMMENDATIONS

The purposes of this chapter are threefold. First, findings are summarized. Second, the chapter describes implications associated with the research questions guiding the study. Since the theoretical framework for the study is mainly interpretive in nature, the chapter does not attempt to make conclusions about the data; rather, it culls nine implications most supported by data collected for the study that may further our understanding of the issues investigated. The third purpose of the chapter is to suggest directions for improved practice in classroom grouping and evaluation, curriculum reform in teacher education, and further research.

Summary

This study examined how 23 experienced fifth and sixth grade teachers in 10 schools in the Coalition for School Improvement: 1) apply criteria for grouping students for Reading instruction; 2) evaluate student progress in classroom reading groups; and, 3) use data from classroom student evaluation in Reading.

Findings show that teachers rely on different criteria to make group placement decisions at three key intervals of the academic year. On the first day of school and before they begin teaching the curriculum, teachers are given lists of students' names pre-assigned to within-class or between-class ability groups for Reading. These lists are determined by the recommendations of the previous year's teachers and the various criteria that administrators and support staff use to create balanced, heterogeneous classrooms. During the first few days and weeks of school, teachers are getting to know their students and at the same time are rearranging their within-class groups from the pre-assigned lists. The data show that the formative evaluation strategies teachers use

to elicit the criteria for modifying grouping arrangements during this period vary depending on whether the reading curriculum was basal or literature-based. Two-thirds of the classrooms in the study were basal-based and one-third was described as literature or novel-based. Five classroom teachers were using combined approaches. After the first month of school, reading groups are usually solidified and teachers' impressions of students have begun to stabilize. From this point on in the year, if group changes occur, teachers in basal classrooms use prescribed basal materials as the primary means for evaluating and regrouping students. Meanwhile, in literature-based classrooms, a broader range of evaluation strategies were used. All teachers cite the results of their professional observations of students as their most oft-used criteria for making group placement decisions.

The relatively new or re-emerging phenomenon of literature-based reading curricula among sampled classrooms confounded, somewhat, comparisons of findings of this research with prior research. Nonetheless, there is general support for these findings from research on teaching that focuses on teachers' preinstructional decisions (preactive, planning), interactive decisions (made during instruction), and, postinstructional decisions (Shavelson and Stern, 1981). This body of research suggests that while teachers make judgments about students' ability and achievement rapidly and rather accurately, the possibility of error in interpreting myriad interactional cues clearly exists. Moreover, once teachers make their judgments of student ability early in the year, assessments of ability usually cease. Some studies indicate that teachers leave these impressions in place and thereafter move on to other forms of assessment that focus more on achievement (Calderhead, 1983). Thus, when teachers "mis-see" what they are observing, the result is often misdiagnosis of student ability and achievement.

The compounded effects for children who suffer misplacement as a result of misdiagnosis are stupendous and have been well documented (e.g., Michaels and Cook-Gamperz, 1979; Heath, 1983). Teachers acknowledge, even express frustration that these errors occur, but tend to see them as mistakes made by other colleagues.

Teachers expressed clear and worthwhile goals for student learning and high expectations that students would succeed in their classrooms. Teachers wanted students to increase their reading and writing skills, enjoy reading, feel good about themselves, and develop a lifelong love of reading. But when asked whether these goals and expectations applied to all the students in all the reading ability groups, less than half of the teachers had similar goals and similar expectations for all their students. The finding that teachers express different expectations for students of different ability has been documented by Good and Brophy (1987), Sorenson and Hallinan (1984), and others whose research also suggests that students in high and low ability groups are exposed to radically different quality and paces of instruction.

Teachers in basal and literature-based classrooms reported using a variety of formative and summative methods to evaluate the progress their students were making in Reading groups. Teachers continuously observed and questioned students to appraise their progress in learning. Section tests were a common summative evaluation measure used in basal classrooms and arts-integrated writing projects were often used as a culminating assessment in literature-based classrooms. Teachers in basal classrooms perceived their evaluation methods to be similar and their performance criteria to be different for students of different ability. Teachers in literature-based classrooms said that their evaluation methods and their performance criteria were different for students of different ability.

Teachers were also concerned with judging a student's progress in relation to where he had started at the beginning of the year rather than in relation to other peers. Teachers said they were concerned about students cognitive, metacognitive and affective development as well. Some teachers may have unintentionally sabotaged these concerns by creating evaluation standards which used exemplary students' work as a basis for judging other students' work.

In the face of a vast array of decision contexts, teachers must act decisively to make classroom evaluation tasks manageable. Data from this research report that teachers store vast amounts of information about a range of characteristics of their students in their mind. Moreover, they use intuition and perception to process evaluative data on a student, almost in an instant. How do teachers deal with the complexity of these assessment demands? There is some indication in the research on teaching of how teachers deal with complexity. For example, Simon (1957) suggested that when faced with an overload of information teachers simplify their view of reality to create a more manageable situation. Teachers find thinking strategies that allow them to process and store information efficiently.

While some studies comment favorably on teachers' skill in correctly judging students' ability and predicting students' performance on standardized tests, other studies suggest that there are numerous problems in teachers' judgments, particularly as they pertain to students of different ability. For example, Brophy and Good (1970) concluded that teachers' expectations of students clearly correlated with differential patterns of interaction between teacher and student. Later, the two authors (1986) verified these results empirically and further concluded that teachers were totally unaware of their differing levels of interchange with students whom they judged to have

different ability. This commonly cited research on teacher expectations and student achievement by Good and Brophy may partially explain why teachers in the present study stated similar goals, high expectations, similar evaluation methods and similar performance criteria for all their students when the referent was the class as a whole. However, when asked specifically about whether similarities or differences existed on these issues, a melange of responses emerged from teachers. This finding implies that the meanings that teachers attribute to generalized qualities of teaching effectiveness have a very local, individual, and classroom-specific context. Teachers do what they think is best for themselves and for their learners at the moment. It may be virtually impossible for standard (positivistic) research paradigms to uncover and make sense of these continually shifting decision nuances and idiosyncracies.

Finally, this study suggests that teachers use student evaluation data primarily for pragmatic purposes such as filling out report cards, justifying grades to parents, and making decisions to keep the classroom running smoothly. Follow-up interview questions stimulated teachers to reflect upon and verbalize a dimension of their professional role that they were unaccustomed to thinking about, revealing notable examples of how they used student evaluation data to assist or improve learning. Another compelling way that teachers said they used evaluation data was to make group placement decisions and to move students to different instructional groups. Findings indicate, however, that for most teachers over the course of the year, such actions were not commonly taken. The responses of teachers further indicate that students were infrequently polled, invited, or taught to participate in the evaluation of their own learning. This evidence contradicts the recognition by many teachers that students are a valuable source of information concerning their own learning needs and interests. Once

again, research on teachers' pedagogical plans, thoughts and decision-making may provide a clue for interpreting this behavior. Apparently, teachers move ahead flexibly, yet inflexibly, once they have planned the content and activities of the curriculum (Zahorik, 1975; Yinger, 1977).

Implications

In the present study, data analysis was based on the self-reports of teachers as garnered through indepth interviews. In complex, interweaving, rapid-paced mental activities such as those associated with teacher judgments and thought processes during interactive teaching, it may be quite difficult to unravel all of the elements involved. However, similar comments and concerns from 23 experienced teachers committed to quality education for children in 10 diverse school settings may provide a sound basis for interpreting findings relating to these matters. This foundation is further strengthened by the conversational, non-antagonistic tone of interviews. Patterns in interview data lead to important implications about student evaluation as conducted by 5th and 6th grade teachers of Reading in selected Coalition schools.

Perhaps the most compelling implication of the study is that while teachers collect a great deal of information about the academic progress, background, and personal characteristics of their students, this data does not translate for teachers into the need to change the practice of ability grouping. On the contrary, grouping by ability remains a pervasive condition.

Teachers acknowledge that there are students in their classrooms who are not doing well in reading either in terms of achievement or on various affective indicators. One might surmise that this would create a reasonable degree of cognitive dissonance for teachers, perhaps leading them to question the utility of grouping students by ability.

Interestingly, teachers participating in this study did not seem to make this interpretation. Teachers were more likely to attribute a host of background and personal characteristics of learners as impeding their motivation and performance in reading rather than the organizational structure of ability grouping. What needs to occur, it would seem, is discovering what impetus needs to be added to alter teacher perceptions toward more reflective, decisive action against the perpetuation of ability grouping as a classroom policy.

This position is taken by this researcher because the debate on the effectiveness and usefulness of ability grouping helps to frame the overall implications of this study. A cumulative body of evidence suggests that grouping children by ability has questionable value for students in higher ability groups and may be detrimental socially and academically to students in lower ability groups. Meta-analyses of the research on ability grouping also suggests that grouping has few short-term effects on achievement, may have a lasting negative effect on the self-esteem of lower-tracked students, may re-segregate children who differ in social class, race, or ethnicity, and may actually lead to increased differences in academic ability and achievement among students. For these reasons, it seemed crucial to gain further insight into why the practice does not seem to bother teachers enough to change it. This is a question for subsequent inquiry that may be more firmly grounded in critical theory.

Second, while evaluation activities proliferate for teachers, evidence suggests that evaluation of students is not a potent force for change in classroom teaching. It remains a largely untapped force, as many teachers feel constrained in giving the careful planning and considerable investment of time that may be required to do it effectively. Nonetheless, teachers are engaged in an astonishing number of evaluative contexts daily.

They struggle to develop evaluation approaches that are fair and sensitive to students, yet may be unaware of how the use of their subjective assessment practices may actually sabotage their well-meaning intents. Thus, classroom evaluation practices of teachers need to begin receiving far more attention than has customarily been the case. The impact of classroom evaluation on students is not often discussed. Therefore, as teachers in the present study alluded to and as Terrence Crooks (1988) has surmised:

> A more professional approach to evaluation would demand the regular and thoughtful analysis by teachers of their personal evaluation practices, greater use of peer review procedures, and considerable attention to the establishment of more consistent progressions of expectations and criteria within and among educational institutions. (p. 467)

The approach that has underpinned this research is that more interpretive studies that involve teachers themselves as colleagues or co-partners in the research process are needed. In this way the questions that teachers themselves view and define as important for increased effectiveness in grouping and evaluation can receive increased attention.

Third, based on interpretation of the view of some of the teachers in the study some classroom reading environments may stimulate student's intrinsic and continuing motivation to learn more readily than others. Intrinsic motivation, defined as a self-sustaining desire to learn, and continuing motivation, defined by Maehr (1976) as a tendency to return to and continue working on tasks away from the instructional context in which they were initially confronted, are related concepts. According to teachers' perceptions, literature-based classrooms may foster more autonomy in choice of learning activities and provide more time for independent and leisure reading the more conventional curriculum approaches. To verify these perceptions, more combined

methods (e.g. empirical and qualitative) research of the benefits of literature-based curriculum approaches and the evaluation of learners needs to be conducted.

Fourth, an important issue discussed in this research was whether student evaluation standards adopted by teachers were comparative-referenced, criterion-referenced, or, based on the effort and improvement of individual students over time. A teacher's choice in this regard may have important implications for the motivation and learning of students of different ability. Comparative-referenced evaluation may hinder students who are in lower groups. Instead, criterion-referenced combined with self-referenced evaluation may motivate higher performance in heterogeneous classrooms. It is impressive that the majority of veteran teachers represented in this study chose the latter evaluation standards for their students in efforts to support and extend their learning.

Fifth, there appears to be a gap between teachers' expressed goals for reading and the evaluation methods they use to determine students' progress toward them. Teachers articulate highly worthwhile goals for students, then seem to get swallowed by the daily work press, their own plans, prescribed curriculum materials, and the momentum of the school year in charting student's progress to the goals. Teachers themselves were unsure of whether students were attaining the goals they set for them. Hence, teachers may need to be encouraged to use evaluation as a more potent tool for determining intellectual potential. Evaluation should not be designed to protect one from the scrutiny of parents and administrators. Teachers said they valued the skills of thinking, writing, and reading for enjoyment, confidence, and transfer. If so, then such important cognitive and affective outcomes[28] must be more deliberately assessed.

Sixth, evidence from teacher perceptions suggests that evaluation and grading are two distinct functions in teaching. While it is reasonable to discover that a teacher's approach to evaluation is largely a subjective, intuitive judgment of progress, students and parents may desire a more "objective" approach to grading, considering the significant cumulative effect that grades have on a student's academic career. In this regard, grading practices of many teachers may be deeply flawed. Unclear or disparate notions of the object of evaluation, the introduction of extraneous factors, the averaging of data on unlike scales, and a general tendency to rate students too high may cause teachers to mismeasure achievement.

Seventh, some teachers make efforts to develop their own assessments, but may lack confidence and training that would enable them to do so for the maximum benefit of students. Teacher-made tests, for example, are the most prevalent form of classroom evaluation at the high school level, yet are virtually non-existent at elementary levels (particularly in reading classrooms where published assessment materials predominate). As a result, the benefits of effective tests as excellent motivational devices could be lost for students. Moreover, a well-designed test can place a rigorous demand on students' recall abilities as well as other higher order thinking processes. There are other useful teaching and learning purposes served by well-designed tests. For example, such instruments can be used to amplify the more important aspects of the curriculum; to stimulate clearer teaching by having items concentrated on the core of what is being taught; and, to provide feedback to the learner about what she/he can and cannot do with respect to the content measured by the test. In short, good classroom tests culled from clear domains of learning, developed in accord with sensible test design principles, and in tune with the intents for learning that teachers have for their learners may be a

teacher's most valid and reliable means for assessing classroom learning. This premise leads to an eighth implication.

Teachers want and need practical help with specific evaluation topics that suit them, their students, and the realities of their classroom. Though they make genuine efforts to do so, most teachers were not trained to meet the rigorous demands of classroom evaluation in Reading. While teachers appear to be fairly accurate in their judgments of students' intellectual ability and achievement, they frequently misdiagnose students' reading problems (Gil, 1980, and Shavelson and Stern, 1981). Moreover, teachers report that they evaluate a broad range of affective and social behaviors of students. Teachers appear to need more assistance in developing more careful means of evaluation for this purpose since it is through these realms that teachers form their most lasting impressions of students.

Finally, an implication is proffered concerning the relationship between instructional grouping, student evaluation, and equal educational opportunity. The structure of ability grouping is largely an unexamined practice in public schools. If the relationship between student evaluation and equality mean different things in different classrooms and schools, then different consequences for children will result. More specifically, if student evaluation is used to support equality of educational *opportunity*, then student evaluation may, like the structure of ability grouping, become just another sifting and sorting process that perpetuates existing inequalities in society. But, if student evaluation supports equality of educational *outcome*, then it may serve as an instrument to combat inequalities in society. Data from this study imply that teachers may be unwittingly contributing to the former of these two issues. It is a situation which teachers must continually monitor as it is difficult to foster the latter ideal in

classrooms segregated by ability. Emancipatory ontologies of research on teaching that address assumptions and concerns of critical theorists are needed to further understanding of these kinds of [hidden] curriculum issues.

Recommendations

Taken together, implications drawn from interview data may be used to infer recommendations for improved practice in classroom pupil evaluation, teacher preparation, curriculum reform, and to suggest directions for further research.

Recommendations for Improved Practice in Classroom Pupil Evaluation

Recommendations for improved practice in classroom evaluation of learners are offered that link findings from research on teacher expectations and classroom assessment environments.

Teacher expectations and balanced criticism. Evidence from research on teaching over the last decade suggests that the expectations teachers have of their students' ability to learn (Rosenthal and Jacobson, 1968; Rist, 1970; Taylor, 1979; Brophy, 1982), the ways teachers group students for instruction (Esposito, 1973; Dreeban and Barr, 1988), and the nature of teacher-pupil classroom interactions (Good and Brophy, 1987) are conditions that have a profound and often lasting effect on students.

For example, Good and Brophy (1987) report that some of the most powerful effects on students stem from academic and non-academic teacher judgments. This could be a positive finding if more teachers were better critics. Unfortunately, evidence suggests that these criticisms may not be as clear and complete as they could be to increase students' attention to the content being learned.

Further, many teachers rely heavily on narrow or limited sources of data, i.e. information registered in gradebooks, to evaluate student learning. Amazingly, teachers use a variety of formative and summative methods to evaluate pupil performance and to assign grades; and yet, these marks that represent these methods in teachers grade books may be subjective and ununiformly recorded (albeit creative) rendering many of the resulting grades invalid. Moreover, it is uncommon to find teachers who are knowledgeable of how to properly select or develop, validate, and use tests for these purposes. Further still, results of observation studies indicate that a significant number of teachers label students according to a preferred set of norms and values, then group and teach them in discriminating ways (Stalling and Hentzell, 1978; Anderson et al. (1979)). Still others use questionable evidence such as class attendance, social behavior, personal appearance, and other extraneous variables in an additive way as primary criteria for determining grades of students.

It is significant that teachers say that they use formative data from a variety of sources for the purpose of connecting with students and discovering alternative ways to engage them in learning. However, pervasive emphasis in some schools on basal activities and pre-packaged content that presumably all learners need to be exposed to, have moved many of these teachers away from the elucidating act of carefully selecting content and objectives that they deem to be most worthwhile for individual students and instead may perpetuate the controversial practice of grouping students by ability for reading.

Improved practice in classroom assessment. Teachers realize that they promote more effective grading practices when they: communicate the criteria for successful performance to students in advance, so that students know what is expected; factor

184

various student characteristics into the grade that belongs there (e.g. achievement) and leave all else out (e.g. attendance, personality, attitude, conduct); use sound achievement data as the basis of grades; keep thorough, appropriate records; and, combine data carefully over time and set appropriate cutoff scores to determine report card grades (Stiggins, 1989).

Recommendations for Improved Teacher Preparation

Many teacher education programs claim the integration of theory and practice as a major tenet of their educational philosophy and have often transformed this principle into an expressed goal. Teachers in this study implied that much of what they learned about student evaluation, they learned "on the job" rather than during teacher training. Meanwhile, debunking teacher education is both a popular and unrelenting (activity) at the present time among educators and lay people alike. There is evidence to suggest, however, that the role of research in teacher education is still being debated. While the debates continue, this author believes that there are at least three important purposes to be served through continued systematic research in teacher education.

First, that efficacy of teacher training to teaching effectiveness has not been established empirically (Evertson, et al., 1985). Neither has it been demonstrated that teacher education is a waste of time as many critics claim. In fact, Haberman (1984) presents a research-based defense of teacher education. Meanwhile, teacher education faculty, teachers in schools, and capable students can begin to conceptualize and conduct modest small N studies together to build a strong research basis for supporting (or rejecting) formal teacher preparation. There is much research needed before the benefits of teacher preparation can be shown to significantly outweigh the benefits of alternative credentialing, or vice versa.[29]

185

Second there is considerable suspicion among the ranks that efforts by education schools to structure pre-service learning may not have desired sustained effects on student teacher behavior (e.g. Locke, 1984). That is, much of what prospective teachers learn in their formal college training is not transferred to their classroom behavior beyond student teaching. Research is needed that provides direction for teacher educators to strengthen the long-term transfer value of the knowledge and skills of pre-service teachers - especially those found to be associated with maximum achievement of children and youth. For example, despite the inclusion of at least one course dealing with classroom management and discipline procedures in most public institutions conferring degrees in education, questions such as: 1) How do pre-service teachers learn to manage classroom routines and discipline the academic and social behavior of children?, 2) How do teachers learn to group students for optimum learning benefit?, 3) How do teachers learn to evaluate pupil learning effectively in a variety of subject matter, learning contexts and for different populations of learners? - are viable candidates for further empirical as well as interpretive research. Other areas of curriculum modification in teacher preparation programs as suggested by this research include: 1) Helping teachers determine the implications of social, personal, and other background characteristics of students for increasing learning; 2) Assisting teachers in designing and validating effective classroom assessments for a variety of grade and subject level requirements; 3) Helping teachers increase sensitivity to evaluation from the student's perspective. Teacher education units can then conduct follow-up studies of their graduates into the first several years of employment to evaluate the sustained effects, both desirable and undesirable, of their pre-service curriculum in these domains.

Third, during the past 20 years there has been a swelling of research activity focused on what has come to be called "teaching effectiveness" research. From the standard research paradigm on teaching these observation studies have filled a perceived technical void in pedagogy by describing specific teacher behaviors that have consequences for how well students score on standardized achievement tests. A persistent problem, however, has been that many teacher educators and school practitioners have not heeded the cautionary language issued by the investigators of these research studies. The authors of this body of research have almost always limited their generalizations of recommended teaching activities to the particular variables and conditions of the study. For example, most of the studies involved a specific population of students (lower ability, lower SES, white (some black and Hispanic 4th and 5th graders)), being instructed in certain subject matter (reading and math), under certain context conditions (public elementary schools emphasizing basic skills curricula where teachers utilize the methods of direction instruction).

A much wider base of research knowledge using alternative theoretical orientations is needed regarding specific teacher practices that are effective with: a broader range of student populations (including increased focus on minorities), across a broader range of subject matter, and under varying contextual conditions. Again, teacher education faculty competent in various theoretical and methodological research orientations should collaborate with their students and local practitioners to increase initiatives that carve further into our understanding of issues of effectiveness in teaching. Taken together, such studies could contribute significantly to the knowledge gains that have already been advanced.

There are certainly other roles for research that can be defined in fostering the improvement of teacher education. Ralph Tyler (1985) advises that historical research be used to investigate the contributions that past studies of teacher education hold for the improvement of teacher education today. Together, the roles suggested provide a constructive framework from which to begin. As Evertson, Hawley and Zlotnik (1985) conclude in their synthesis, "Making a Difference in Educational Quality through Teacher Education:"

> ...Hundreds of studies related to teacher education are available, but the lessons they teach do not add up to a particular model for improvement around which teacher educators should rally. Rather, this seems to be an opportune time for experimentation - and evaluation. (p. 9)

Will teacher educators seize the opportunity for improvement now begging and lead the profession forward under a balanced, self-imposed mandate that lasts? Or, will this important group of educators continue retreating helter-skelter, fearing mandates imposed from decision-making bodies outside the profession, and trying weakly to make required curriculum changes - changes that can only last until the next cycle of reform comes around?

Recommendations for Curriculum Reform in Public Education

The resurgent cry of the 1990s will continue to be that public education must be improved. Educational leadership, school environments, teaching processes and learning outcomes must be improved if our nation is to remain competitive in a rapidly changing world. Of the need for improving American public education, few will disagree. As in the past, however, the debate will rage when the issue turns to specifics.

Improvement is a cultural value. Cultural values are bound to a particular system of beliefs, assumptions, philosophy, and ways of knowing and behaving that

together comprise the essential elements of a culture. Though improvement may be an agreeable standard in our shared American culture at the broader societal level, it becomes a vague, relative standard when considered and compared at other levels within the society. Improvement is a "fuzzy" concept indeed when debated without careful attention given to the myriad purposes of public schooling and the pluralism of values reflected there. Therefore, each community, each institution, and each individual with a vested interest in education must begin to spell out and make relevant what it means to improve public education in America.

Improvement of public education that is lasting must come from the "inside-out," from those closest to the learner to those most remote. The heart of education, the teaching-learning process as it emerges day-to-day in classroom learning environments across this country, can begin to be improved by helping teachers sharpen their own interpretive skills in evaluation. This has radical implications for traditional curriculum paradigms in teacher education. Rather than being peripheral to research in teaching, teachers will more and more be invited as equal partners or even expected to lead the research enterprise (Florio and Walsh, 1980; Bolster, 1983). This means that the prescriptive approaches that characterize many teacher education programs as well as state and school district curricula, may be transformed to include more emphasis on the situated meaning perspectives of teachers and students concerning the daily phenomena of classroom life. This is a promising direction in curriculum reform which is supported by this author.

Recommendations for Further Research

Five recommendations for further research are advanced.

Minority learning needs. Teachers need to appraise curriculum materials critically to determine their relevance (or bias) and educational value for specific groups of learners (e.g. black students, Hispanic students, Asian students, female students, economically poor students, physically handicapped students, etc.). Few teachers, indeed, seem aware of how to use available cultural and physical evidence constructively as a basis for acknowledging the personal integrity and intellectual capacity of black and Hispanic students (Nobles, 1974; Gwaltney, 1980; Hale, 1982). As a result, the American education system has largely failed in meeting the needs and aspirations of significant numbers of its children-clientele. More focused research is needed on the implications of roots, culture, and learning style in developing curriculum and teaching methods for successfully educating all children, but particularly low achieving poor children and increasing numbers of black and Hispanic children. Further research that focuses on grouping and evaluation of minority learners with a broad range of demographic characteristics is needed. Also, more case studies and rich ethnographies of individual minority children, their families, and communities (e.g. Nieto, forthcoming) should be undertaken. The research orientations for this important work need to continue moving away from positivist paradigms and rejoin work from more interpretive and critical perspectives.

Richer, fresher concepts of evaluation. Another recommendation is that constructive conceptualizations of evaluation continue to be explored and internalized by the teacher-evaluator. From a broad understanding of various assumptions, concepts, methodology, and value orientations in current use in the field of evaluation research, teachers can adapt and purposefully blend different approaches to obtain useful information for instructional decision-making. The conceptual ideas of Eisner (1985) are

particularly promising for the classroom teacher-evaluator. Further cultivation of skills in connoisseurship and criticism should be particularly appealing to the teacher who regularly evaluates instruction and uses the information to improve teaching and learning. Similarly, the process of adversary evaluation, developed by Wolf (1975), can be adapted to the reading classroom, challenging teacher and pupils to join in solving classroom problems after considering viable alternatives. In addition, further inquiry into the meaning and practice of evaluation from the viewpoint of teachers needs to be conducted using greater methods of triangulation. Data gathering strategies may include classroom observation, microethnography, extended interviews and fieldwork.

Student self-efficacy. Providing learning environments that foster self efficacy in students may be a powerful way that teachers can help their students to persist with more difficult learning tasks. The concept, defined by Bandura (1977), has strong implications for learners at both ends of the achievement continuum. In a descriptive study on student evaluation, an experienced classroom teacher frames the case:

> Very often students will start out wanting to get bits of approval halfway through something. They ask, "Does this sound all right?" I try to get them to see that it is their piece of work and they should work on it until they feel good about it and then bring it to me to compare their perspective with my perspective. Self-evaluation is a form of self-discipline. It is a chance to look at what you have done, to feel a sense of achievement in what you've done and to be able to look at your own work without being defensive about it. These are important attributes for kids to develop. (Thiessen and Moorhead, 1985, p. 5)

Results of this study indicate that teachers are not accustomed to involving students in evaluating their own learning. More interpretive studies that focus on the views that students have of their own learning and specific conditions that would help them learn better are needed.

Follow-up studies on classroom evaluation. It is worthwhile to conduct follow-up studies that extend the methods of data collection from the present study. As qualitative research methods continue to be refined, more descriptive studies on classroom evaluation environments are needed that further contribute to a comprehensive understanding of teacher practices in pupil evaluation. The use of document analysis, stimulated recall, journal keeping, classroom observation and video taping, and interviews with students are recommended approaches.

More critical studies that help to empower teachers. Given the overall findings and frame of reference for the present study, it is clear that more studies are needed that further our understanding of how knowledge leads to change and how teachers can be fully empowered to act in ways that they believe benefit learners. Teacher preparation programs and reformed political structures of schools are promising foci for such inquiry. Research studies that take a critical perspective on these issues can help inform constructive action. This study provides a documented basis for the need for further inquiry in this direction.

Closing

This study has attempted to provide a conceptual framework and suggest directions to assist teachers in thinking about and resolving practical problems in the evaluation of student learning. Evaluation is an essential yet often clouded dimension of curriculum. While teacher activities in pupil evaluation are prolific, more studies are needed that help to clarify the complex issues associated with this crucial function of teaching.

A major theme has been that teachers, reflecting on their daily practice, are in the nearest position to alter conditions that will directly result in better learning for

children. To complement this critical undertaking, newer and more sensitive approaches will need to be cultivated to assist teachers in evaluating the full range of accomplishments for diverse student populations. Teachers also need to develop deeper critical consciousness and be empowered to take action for constructive change.

The challenge of educating all of the nation's youth has never before been so complex, so difficult. Yet, it is a challenge that forward looking educators must meet. Competent, confident teachers who constructively evaluate student learning while at the same time maintaining a healthy tension against learning conditions that may be hindering the growth of learners, will help us to know whether we are meeting this challenge successfully in the years to come.

APPENDICES

APPENDIX A

LETTER OF INVITATION TO PRINCIPALS AND SUPERINTENDENTS

February 1, 1989

Dear Principal,

The Coalition for School Improvement is continuing its systematic inquiry into conditions in schools which promote equal and high quality learning for all students. We are presently conducting a study to learn more about the complexities of pupil evaluation in elementary classrooms, particularly as it is carried on by upper elementary teachers who use some form of ability grouping to teach reading. The purpose of this letter is to invite your school to participate in this important study.

Everyday, elementary public school teachers evaluate the learning of their students. Further, at the end of a marking period teachers have the taxing responsibility of assigning a letter or numerical grade to each child in a variety of subjects. These reported grades symbolize a child's progress in school. Hence, the tasks of evaluating and grading children's learning are not frivolous. These tasks are perhaps even more demanding for teachers during reading instruction since it is here that teachers are often managing three or more groups simultaneously and are often evaluating and using data for making instructional decisions. Unfortunately, too little is known about how elementary teachers carry on the processes of evaluation in the classroom.

The purpose of this research is to examine perceptions teachers have of their student evaluation practices during reading instruction. The criteria teachers use to group students for instruction, the means teachers use to evaluate progress for individuals in the various groups, and the decisions teachers make for using the information they obtain from their student evaluations are three important parts of this study. To provide data for the study, we are interested in interviewing experienced 5th and 6th grade teachers of reading in coalition schools.

When our research is completed, we will provide the participating school with a concise report of our findings. This may be used by the school staff to reflect upon their present practices and determine future priorities for curriculum and instruction.

Thank you for considering this research. We will follow up with a telephone call to you in two weeks to learn of your decision and, if you agree, to schedule a meeting with your 5th and 6th grade teachers to provide more detail about the study. If you

196

have any questions about this research, please feel free to contact Kriner Cash at the number given below.

Best wishes for a meaningful and productive new year.

Sincerely,

Robert L. Sinclair,
Director and Professor

Kriner Cash,
Principal Investigator
    and Chair, Education Department
    North Adams State Colleges, MA
(413) 664-4511 ext. 381

cc: Superintendent

APPENDIX B

LETTER OF INVITATION AND WRITTEN CONSENT FOR TEACHERS

APPENDIX B: Letter of Invitation and
Written Consent for Teachers

Written Consent Form

PERCEPTIONS OF SELECTED UPPER ELEMENTARY PUBLIC SCHOOL TEACHERS TOWARD GROUPING AND EVALUATION OF PUPIL LEARNING PROGRESS IN READING

Dear Colleague:

I am Kriner Cash, a graduate student at the University of Massachusetts in Amherst. I am also Assistant Professor and Chair of the Education Department at North Adams State College in North Adams, MA. My mentor, Robert Sinclair, Director of the Coalition for School Improvement, and I are conducting a study to learn more about the complexities of pupil evaluation in elementary classrooms. The purpose of my doctoral research is to investigate and describe the perceptions teachers have of their pupil evaluation practices in instructional groups for Reading. I am especially seeking 5th and 6th grade teachers of reading in member and affiliated coalition schools in Western Massachusetts who use some form of ability grouping for instructing their pupils in reading. If you agree, you will be one of approximately twenty participating teachers.

As a part of this study, you are being asked to participate in one in-depth interview. The interview will be centered around three open-ended questions. The first question will focus on learning about the various criteria that you have used to assign or place your students into grouping arrangements within the classroom. Second, since you are now well into the academic year in reading instruction with the students in your classroom, we will also be interested in better understanding the ways that you are evaluating the progress that your students are making within the groups. The third question will focus on learning more about the kinds of instructional decisions you make after you have evaluated your students' reading performance for a unit, section, or full term of study.

The interview questions have been piloted with a small sample of teachers like yourself in order to make it possible to obtain vital information while requiring a minimum amount of your time. The total time of the interview should not exceed 45 minutes.

My goal is to analyze the materials from your interviews in order to understand better your practices in pupil evaluation and those of other upper elementary reading teachers. As part of the dissertation, I may compose the materials from your interviews as a "profile" in your own words. I may also wish to use some of the interview material for journal articles or presentations to interested groups, or for instructional purposes in my role as a teacher educator. Existing studies suggest that there is a great deal more to

learn and share about student evaluation if educators are to help each other feel more confident about this part of their role.

Each interview will be audiotaped and later transcribed by me or by a typist (who will be connected with your school and who will be committed, as I am, to confidentiality). In all written materials and oral presentations in which I might use materials from your interview, I will use neither your name, names of people close to you, nor the name of your school or town. Transcripts will be typed with numerical codes for names, and in final form the interview material will not refer to any teacher by name or school.

You may at any time withdraw from the interview process. You may withdraw your consent to have specific excerpts used, if you notify me at the end of the interview series. If I were to want to use any materials in any way not consistent with what is stated above, I would ask for your additional written consent.

Finally, I wish to express my gratitude to you for your willingness to participate in this study. Your experience and wisdom are significant ingredients for the study. We are proud of our special focus in the Coalition of trying to determine together ways to increase learning for all students. This study is designed to provide further knowledge for this important objective. We look forward to sharing the results with you in the near future.

_____

I, _____, have read the above statement and agree to participate as an interviewee under the conditions stated above.

_____
Signature of participant

_____   _____
Signature of interviewer   Date

APPENDIX C

PRELIMINARY DATA SHEET

CLASSROOM PUPIL EVALUATION IN READING:
PERCEPTIONS OF 5TH AND 6TH GRADE PUBLIC SCHOOL TEACHERS

PRELIMINARY DATA

I.   TEACHING EXPERIENCE
Years Teaching Experience: Grades Taught:

    3-5   6-10   11-19   20+   K   1   2   3   4   5   6   7   8   9   10   11   12

Years Teaching Reading: At this Grade:

    3-5   6-10   11-19   20+   0-2   3-5   6-10   11-19   20+

II.   EDUCATION
Highest Degree Attained:
    B.S.   B.A.   B.A. + 15-30   M.A.   M.Ed.   Ph.D.   Ed.D.

III.   SEX
    M____   F____ .

IV.   CLASSROOM STRUCTURE/ORGANIZATION

    ___Self-Contained   ___Departmentalized   ___Team Teaching
    __single grade   __single grade
    __combined grades   __combined grades
    ___Other (please describe)

V.   CLASSROOM GROUPING AND INSTRUCTION FOR READING
(Please check all that apply)

Grouping Arrangements: Instructional Approaches:
___Whole Class activities   ___Basal Series
___Small Group (varied ability)   ___Whole Language
___Small Group (similar ability)   ___Literature/Novel
___Pairs or teams of three   ___Subject matter
___Peer Tutors   Integrated___
___Individualized   ___Other (please describe)
___Ad Hoc
___Continuous
___Other (please describe)

APPENDIX D

INTERVIEW GUIDE

CLASSROOM PUPIL EVALUATION IN READING:
PERCEPTIONS OF 5TH AND 6TH GRADE PUBLIC SCHOOL TEACHERS

The purpose of this interview is to learn more about how you evaluate the progress that your students are making in Reading. Three major questions have been developed to provide information for this purpose. Several follow-up questions are asked to permit you the opportunity to more fully address the potential complexity of each main question.

If you should wish to add more information to your response after the interview has been completed, a reprint of the three major interview questions is provided for you in a pre-stamped envelope. You are invited to reflect upon the questions further and mail in your additional responses to me within ten days. Thank you.

1. · What criteria do you use to group your students for Reading instruction?

2.   Are there similarities and differences in the ways that you evaluate the progress your students are making in the Reading groups? If so, what are they?

Similarities

Differences

3.   When you have evaluated the progress your students are making in Reading, what do you do with this information? How do you use it?

Once again, thank you for your cooperation. Your generosity and your candid responses are most appreciated. Please return your question sheets within ten (10) days to:

Kriner Cash, Chair
Education Department
North Adams State College
North Adams, MA 01247
(413) 664-4511 ext. 381

Description of how Teacher keeps records of student progress, marks student work, grades student performance, and other student evaluation practices (optional):

Teacher Record Book

207

Student Work Samples

Other

1a. What kinds of assessments do you use to assign students to reading groups within your classroom?

b. I am comfortable with the assessment strategies I use to assign students to reading groups within my classroom.

<div align="center">

SA   A   U   D   SD

</div>

additional comments:

c. Of the types of assessments that you use which would you say you rely on most to make group assignments in Reading?

d. Do you have goals and expectations for your Reading program? (Y   N)

e. If so, can you describe them?

<u>Goals</u>   <u>Expectations</u>

f. Are these goals similar or different for each of your reading groups? (Y   N) How are they similar or different? Please give examples.

<u>High</u>   <u>Middle</u>   <u>Low</u>

<u>High</u>   <u>Middle</u>   <u>Low</u>

2a.   How many (ability) groups do you have in your Reading classroom? ____

b.   What specific methods do you use to evaluate student learning progress in the reading group(s)?   What performance criteria do you use to judge student work?

        i.e.   When you are evaluating your students, how do you do it?

<u>Methods</u>   <u>Performance Criteria</u>

<u>High</u>

<u>Low</u>

c.   Are your evaluation methods and performance criteria similar or different across the reading groups?

        i.e.   Let's consider the low group for example.   Is there anything unique to the way you evaluate the learners in this particular group?   The high group?   Middle group, etc.
            OR
        i.e.   Is there anything you do to tailor your evaluation practices for a particular individual or group?

<u>Similarities</u>   <u>Differences</u>

d.   Would you describe your approach to pupil evaluation in Reading as primarily comparative-referenced or criterion-referenced?   (Briefly define the two approaches.)

e.   What kinds of information do your evaluation methods yield?

3.a. What are the ways that you use data from pupil evaluation in Reading?

(Turn off tape)

b. How many children have you moved from one group to another this year?

How many children are in the class?

c. In what direction have the children moved - e.g. low to mid, mid to high vice versa, or, out of the classroom altogether - e.g. referrals for core testing, special education, resource room, etc.?

d. Are there other determinations you make from the data you collect about student learning?

e. What factors help and what factors hinder your ability to evaluate student progress in reading effectively?

Helping Factors   Hindering Factors

Thank you...

(Turn off tape)

APPENDIX E

THANK YOU LETTER TO TEACHERS PARTICIPATING IN THE STUDY

June 1, 1989

Dear

Thank you for your gracious cooperation in permitting me to interview you during the period of March 20 - May 15 for the study on pupil evaluation practices in Reading. The data collection phase of this coalition study has been completed. In all, twenty-three experienced 5th and 6th grade teachers in ten schools participated. You shared your perceptions as well as selected projects and record-keeping documents about how well students in different ability groups are succeeding in your reading classroom.

The interviews have produced material which I am sure you will find fascinating as you continue to reflect upon your practices in student evaluation. Moreover, I have learned valuable lessons from each of you. Observing the unique character of each school and having the opportunity to absorb a part of your teaching philosophy has been especially exciting to me. I now turn to the challenge of organizing and interpreting the large quantity of information in a way that can be useful to you and pre-service teachers as well.

Once again, thank you for your time and thoughtful responses. I will send you a report as soon as my study is completed. Best wishes for a successful close to your school year.

Sincerely,

Kriner Cash
Principal Investigator

Robert L. Sinclair
Coalition Director and Professor

cc: Principal

APPENDIX F

ANALYSIS OF CLASSROOM STRUCTURE, GROUPING ARRANGEMENTS,
AND INSTRUCTIONAL APPROACHES OF THE TEACHERS
PARTICIPATING IN THE STUDY

To get an idea of the diversity of classrooms, information gathered from the
Preliminary Data Sheet (see Appendix C) regarding classroom structure, grouping
arrangements, and instructional approaches of the 23 teachers participating in the study
was analyzed. Results indicate that a unitary pattern did not emerge for sampled
reading classrooms; rather, a mixture was being employed by teachers on all three
variables.

Classroom structure. The school sample involved five member and five affiliates
of the Coalition for School Improvement. The classroom structure of teachers in
member schools can be described in two categories - self-contained (N=5) and
departmentalized (N=6). In the affiliate schools, there were seven self-contained and
four departmentalized classrooms, and one team taught, combined-grades classroom. In
departmentalized grades, teachers typically taught their own reading classes and students
were homogeneously assigned to one of two or more between-class ability groups.
Table 11 describes the classroom structure for Reading of teachers in the study.

Classroom grouping arrangements. It was anticipated in the present research
that teachers would be using some form of ability grouping for instruction in Reading.
Interview data from this study generally support this premise. Eighteen of 23 teachers
report the use of within-class ability groups for Reading. Thirteen teachers report
having two (N=7) or three (N=6) ongoing ability groups. These groups are described
on a continuum as comprising of students of High, Low or High, Middle, and Low
reading ability. Two teachers report having four groups, and one teacher's classroom

Table 11

Classroom Structure/Organization for Reading
of Teachers in the Study

School (N=10)	Classroom Structure	Grade	Comment
MEMBER			
A	Self-contained	5	
	Self-contained	6	
B	Departmentalized	5	Teachers teach own Reading
	Departmentalized	5	class; between-class ability
	Departmentalized		grouping used; open space classrooms
C	Self-contained	5	Open space classrooms
	Self-contained	5	
D	Departmentalized	5	Teachers teach own Reading
	Departmentalized	6	class
E	Self-contained	5	
	Departmentalized	6	Between-class ability grouping used
AFFILIATE			
F	Self-contained	5	Low-ability within-class group
	Self-contained	5	goes to Reading Spec.
	Departmentalized	6	Teachers teach own Reading
	Departmentalized	6	class
G	Self-contained	5	
	Self-contained	6	
	Self-contained	6	
H	Departmentalized	6	Between class ability grouping
	Departmentalized	6	used
I	Self-contained	5	
J	Self-contained	5	Special ed. students in grade 5
	Team Teaching	5-6	group

groups fluctuate from three to six. One teacher reports using a whole class individualized approach exclusively.

Six teachers had homogeneous class-assigned ability groups. Two of these teachers further divided their Reading class into three and four within-class ability groups, respectively. Five of 23 teachers said they were not employing within-class groupings. Four of these teachers, however, had responsibility for a homogeneous, class-assigned group of students labelled as Low, Middle, or High ability in Reading.

Further, while small, continuous groups comprised of students of similar reading ability was the predominant grouping arrangement in the sample, teachers also reported the use of a variety of other grouping arrangements including frequent use of whole class activities, efforts to individualize reading instruction, and ad hoc groups to re-teach a particular concept or skill. Several grouping arrangements recommended in the current literature as contributing to increased learning, such as placing students in small groups of varied ability for cooperative learning purposes, in pairs or teams of three for a common purpose, or grouping them for peer tutoring, were not common practices of the experienced teachers in this study.

Instructional Approaches. The use of basal reading systems were identified in 15 of the 23 respondents' classrooms. Thirteen teachers reported a literature-based approach being used for Reading. At least two teachers were trying to combine novel-based and basal approaches and about a third of the teachers said they were making notable efforts to better integrate reading instruction with other subjects learned in school. Whole language approaches were not widely reported (N=1) for the middle grades studied in this research. Table 12 presents a grid of classroom grouping arrangements and instructional approaches for Reading reported by teachers participating in the study sample.

216

Table 12

Grid of Classroom Grouping Arrangements and Instructional Approaches for Reading Reported by Teachers in the Study

Classroom Grouping Arrangements:	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Whole Class activities		X	X	X	X	X				X	X	X	X					X	X	X			
Whole Class similiar ability					X	X				X	X					X	X						
Sm. Group (varied ability)																							
Sm. Group (similar ability)	X	X		X				X	X			X	X	X	X	X	X	X	X	X	X	X	X
Pairs or teams of three																							
Peer Tutors								X													X	X	
Individualized	X				X	X				X	X									X			
Ad Hoc	X	X	X	X	X	X				X						X	X						
Continuous	X	X			X	X	X				X	X	X	X	X			X					

Continued next page

217

Table 12 cont.

Classroom Grouping

Arrangements:

<u>Teachers (N = 23)</u>

	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Instructional Approaches:																							
Basal Series			X	X			X	X	X		X	X	X	X	X	X	X	X	X	X	X		
Whole Language												X											
Literature/Novel-Based	X	X			X	X				X	X	X				X	X		X	X	X		
Subject Matter Integrated	X		X	X																X	X		
Other																							

218

APPENDIX G

EXPRESSED GOALS FOR READING OF THE TEACHERS
PARTICIPATING IN THE STUDY

Interpretation of Teachers' Expressed Goals for Reading

All 23 teachers participating in the study said, yes, they had goals for their reading program. When probed further, most (20/23) provided clear, articulate statements of their goals and expectations. Collectively, teachers expressed six kinds of goals stated in learner terms. These goal categories are listed in order of frequency with which they appeared in the data: 1) Enjoyment; 2) Skills development; 3) Transfer; 4) Confidence; 5) Writing; and, 6) Subject matter integration. Eighty-four goal statements were made in response to follow-up question in the interview. Only two goal statements were stated in terms of what teachers themselves wished to do - i.e. "I would like to use novels in the classroom more," "I would like to be able to discern better when a faster or slower pace is needed for the particular story or tradebook we are reading." To facilitate interpretation of the findings for this important follow-up question, a plausible link between teachers' stated goal choices for students and basic principles culled from research on reading and learning are briefly presented next along with selected teacher quotes for each category.

Enjoyment. Certainly, for reading to have sustained effects value for children and adults, there needs to be some measure of enjoyment and personal satisfaction derived from doing it. Noted educators have consistently confirmed the importance of interest and effort (Dewey, 1913) and motivation (Tyler, 1949) as essential conditions for conscious human learning. These principles apply forcefully to the act of reading. Teachers are clear about their desire for children to feel as they do about reading. They want children to have increased motivation, interest, love and appreciation for literature.

> I think that they enjoy reading,...I think that's an important thing. I think it is very important in terms of teaching Reading that you motivate...get

the kids interested in what they're reading...if they're not interested then they're not going to do their best, they're not going to try.

Have an appreciation for reading. I want them to enjoy reading...that it's something that's alive...something that changes.

That students become invested with a personal excitement about the reading process, that they become engaged in it and it is not a passive activity. To that extent, the more commitment and time that one puts into their reading is something I'd like to see.

My goal this year is to have the students enjoy reading, where reading becomes a pleasurable experience and not a forced experience.

My main thrust in my room is to turn kids on to reading and to read good literature...my main thrust for my 5th grade class is to turn them on to reading so they'll enjoy the reading and want to read.

I would like them to have joy connected with reading. I read constantly. I read children's books all the time. I read grown up books. I just adore it, and it's such a huge part of my life that, if anything, I would like to communicate that excitement to the kids. If they don't feel it themselves, to at least acknowledge that it does exist and for them to start to question themselves, "Gee, she gets so much out of it, she's so enthused and so excited about this, and there's other kids in the class who feel this way too...what are we missing?" I want them to feel as though everyday, when I pick up a book, I'm having a party. I want them to be part of that. I'm trying to create a desire in them to love literature.

Skills. "Reading is the process of constructing meaning from written texts. It is a complex skill requiring the coordination of a number of interrelated sources of information" (Report of the Commission on Reading, 1985, p. 7). Reading is a continuously developing skill, improved through practice and through a proper balance between practice of the parts and practice of the whole. The process begins with a child's earliest exposure to the printed word and a literate culture and continues throughout life (Chall, 1983). Teachers understand the importance of skill development and expressed increased fluency, word attack, vocabulary, comprehension, and critical interpretation skills as a major goal for their students.

As far as skills, my whole interest at the 6th grade level is more to the critical thinking skills - cause and effect, summarizing, recognizing bias, interpreting and drawing conclusions - various levels in those areas, as opposed to decoding structural analysis which are dealt with, but not as my major goals.

And then there's the obvious goal that, yes, I expect their vocabulary level to develop. The vocabulary at this level is usually way above what the comprehension skills are, so another goal is that they be able to read a simple short story on their own, and be able to understand and interpret it without always the input and spoon feeding from me.

The other thing I would like them to do is read with deeper comprehension...whatever they read, for whatever reason, even if they don't have that love and that joy. Can they understand it better? Can they get a grip on characters? Do they understand the plot? Do they understand theme? That a real big thing with me...what's the theme of the story? What lesson can be learned from it that applies to them?

The goals are primarily...the first level would be that they be able to read. That means that to every degree they work on their reading skills...I provide that as much as I can.

Transfer. Perhaps one of the most crucial and most neglected principles of

teaching and learning today is the principle of transfer. What is learned in school may

seldom be practiced or reinforced elsewhere. Children need frequent and multiple

opportunities to practice reading in contexts other than classrooms if they are to become

adept readers and thinkers. Teachers appreciate this need and are articulate in their

expression of transfer of learning as a significant goal for their students.

> I want a kid to be able to think, to see clearly, and be able to make a solid acceptable choice. And, I think Reading is the place to teach that kind of thing, so that later on in life...either a political decision or a personal decision...that capability will be there. It takes a lot of years to get to that decision, and it's important that they comprehend what their choices are...Later on, I want them to be able to make the right moral choice, ethical choice, and to have integrity. And also, not to be used by people who are more powerful than they are. I want them to see their choices clearly and that, I think should start now. Starting in the lst and 2nd grade and working their way up through all the Reading programs. It's the bigger picture.

I want them to develop the habit of reading as well. I want them to see reading as a real thing, not just as a subject in school. I know it's a constant battle with the video bombardment to get people to take the time to read, but I want children to realize over a period of time that reading is a very personal thing, too. In the present culture, a lot of kids come from difficult family situations and reading can be a way of getting away from all that. You don't like to use reading as an escape, but I prefer to think of it as a plus, as going somewhere else, or being somewhere else in a different word and escaping from the one that you're in.

It makes me feel good to see a student who has been a reluctant reader pick up a book and start reading in his/her spare time. I think maybe that would be the main goal, to get kids really wanting to read and realizing that it can be enjoyable as well as informational.

Confidence. Confidence to attempt and successfully engage in a learning task is an essential condition for human learning (Tyler, 1988). Many children would like to improve their reading but fear that they may be ridiculed and perceived as failures by their peers. Teachers say they want their students to feel good about themselves, to have confidence in themselves and in their reading, and for low achievers in particular, to be comfortable with oral reading.

I would like them to be more comfortable with reading novels, literature, be able to talk about it and especially, write about it. I would like them to feel good about themselves and successful in the work that they've done...I would like them to be able to feel that they've grown, and I try to point that out to them as often as I can.

I want them to feel comfortable. Some of them really freeze up. A lot of kids point. I want them to read with their eyes. They're very nervous at first, but as the year goes on, one of my main goals is that they get rid of that fingerpointing and let their eyes do the reading. If you're comfortable, you'll ask questions if you don't know what a word means. You won't feel embarrassed.

I think my whole point is, I would like the student to leave the room confident in knowing what he can do and, socially being confident about himself...doesn't expect to have to do everything perfectly...If you fall flat on your face...so what, we all do, but you get up and you do the best you can...and, if you do that most of the time, that's good enough.

Writing. Like reading, increased skill in writing comes from increased practice and well-guided instruction. Data from National Assessment of Educational Progress studies over the last decade have consistently shown a positive correlation between these important skills (NAEP 1984, 1989). Better readers make better writers. Teachers themselves have observed this relationship in their classrooms and some have set as a goal, increased and diverse writing opportunities for their students.

> Pretty soon we're going to be starting Mythology...the students will be reading a variety of myths and doing different writing activities with those.

> We read mystery stories one month and we wrote mystery stories. They were fantastic. It motivated me to try to have more chances for kids to write in creative ways...for instance, they had to make their book a mini-series, advertise it, and make it into a TV Guide.

> In my classroom work often involves writing about their understanding about things that have happened, and what they've read. One flaw is that it [assessment for ability-grouping] depends a great deal on their writing ability as well...

> Sometimes we'll compare...We'll use video and books...comparing and contrasting...A lot of writing, too. That's one of the aspects that I've gotten into this year...A tremendous amount of writing, along with reading.

Subject matter integration. The idea that reading instruction and subject matter instruction should be integrated is an old one in education. (1895 NEA proceedings) More recently, the 1985 Report of the Commission on Reading suggests that "as proficiency develops reading should be thought of not so much as a separate subject in school but as integral to learning literature, social studies, and science." (p. 61) Teachers in enriched classrooms recognize this important interdisciplinary connection and have made provisions for their students to recognize it as well.

> Lots of times we'll use another Social Studies book. My Reading groups have just finished three weeks reading about France in a 6th grade Social

Studies book, because we did a unit on France and we went to the Ritz Carleton to meet the chefs from France...we used magazine articles, we used Time Life Cookbook from France...We used the French/English dictionary...

We're going to be starting a study on aquatic life...we're going to Mystic Seaport and we're adopting a sea lion and that's as a group project...whether you're reading high school level or 3rd grade level, it will be done as a joint effort.

Throughout the year, periodically, we have "units" of thematic literature. For instance, we did Shakespeare during the month of March and we did more directive reading then. Each of us had different plays that we read with the kids on different kinds of activities about Shakespeare, but it was a variety of the plays that were read...Poetry, we've done...I have decided that given the students that I have, and their interest level and ability level, I was going to have them read science fiction...They've been reading science fiction for the past three weeks.

In sum, teachers expressed clear goals for their reading curriculum related to their students' intellectual, social, and personal development. Teachers wanted their learners to read for enjoyment, skill enhancement, and transfer, but were not always confident of the progress they were making toward these goals with individuals or small groups.

END NOTES

1. John Goodlad (1966) proposed a conceptual framework for curriculum development that suggested five levels of curriculum decision-making and their interrelationships within society - ideological, societal, institutional, instructional, and personal.

2. Jackson (1968), describing the way teaching is in his seminal study of its kind, points out that the professional time of teachers actually occurs in three modes - public, semi-private, and private. Although Shavelson (1983) and others have actively cultivated a new field of research that seeks to inquire simultaneously into all three modes, little yet is known about the instructional decisions that are made during the semi-private and private episodes of teaching.

3. There is mounting research evidence to support these uncomfortable feelings of teachers in the area of pupil evaluation (e.g., Zahorik, 1975; Gil, 1980; Shavelson & Stern, 1981; Barnes, 1985; Stiggins & Bridgeford, 1985).

4. Investigation of the major concepts under study - classroom pupil evaluation and teacher decision-making in ability groups for Reading - could proceed from various analytical and methodological stances. For example, one could study the problem from linguistic, historical, critical, interpretive and/or normative analytical frameworks; further, one could employ techniques of descriptive, historical, correlational, experimental, ethnographic, and ex post facto research to the study being proposed.

5. Congruent with the definition by Glaser, norm-referenced assessments can also be referred to as "comparative-referenced." This terminology is used in interviews with teachers to reduce confusing jargon. Comparative reference is also used in the discussion of findings in Chapter IV.

6. This latter pattern, within-class ability grouping, is done routinely for reading instruction in elementary schooling (Hiebert, 1983; Cazden, 1985), and is the focal instructional grouping arrangement being investigated in this study.

7. Recent research by Sinclair and Ghory (1987) on marginal learners provides an important conceptual framework for better understanding and responding to the needs of this significant population of students.

8. As described in the section, Meaning of Terms, there are, however, important conceptual distinctions among these and related terms.

9. Source: Department of Education, National Center for Education Statistics, Washington, D.C.: U.S. Government Printing Office, 1987.

10. Scrivens is explaining the problems he encountered with goal-based evaluation as he worked on a committee commissioned by the United States Office of Education to evaluate disseminable products of regional laboratories and R & D centers throughout the country during the late 1960s.

11. The author of the present study believes that the ideas put forth by Eisner are significant ones and may have shared meaning for many teachers in reflecting upon their own notions and theories of classroom evaluation.

12. See John I. Goodlad, et al., 1975 and Ralph W. Tyler, et al., 1981 for examples of such reports for the School of Education at the University of Massachusetts, Amherst.

13. This definition of evaluation of instruction represents a synthesis from the work of a number of curriculum scholars.

14. These periods of inquiry represent largely western (European) cultural constructs that exclude from consideration vital alternative orientations for deriving knowledge about the world that stem from, for example, African, Asian, Middle Eastern, and cultures below the equator.

15. Gage also argues that most of the controversy and tension over alternative paradigms in research is because most scholars prefer to think and pursue lines of inquiry from the theoretical orientation in which they were trained (i.e., in graduate school). This argument has a plausible ring to it and is similar to the idea put forth earlier about evaluators using methodological approaches with which they are most familiar.

16. For an historical and more detailed description of the work of these early sociologists, see A. F. Wells (1939), "Social Surveys." In F. C. Bartlett, et al. (Eds.), The Study of Society, London: Kegan Paul and Co.

17. The relationship between knowledge and reality (truth) that perhaps many educators, researchers, and the "average citizen" rely on is that of the pragmatist. John Dewey, of course, was a leading spokesman for this position, which essentially holds that all knowledge is produced by human beings and that we can never distinguish between knowledge and truth. If something works in practice it is true (or assumed to be so). A truth that is not supported by further empirical study is modified or discarded.

18. Judith Green and her colleagues and students at Ohio State University have been advancing theory driven research that applies concepts of symbolic interactionism, cognitive anthropology, sociolinguistics, and ethnography to the systematic study of classroom culture. (e.g. Green, 1983; Puro and Bloome, 1987)

19. The substance of the policy recommendations by Meier et al. are faithfully reproduced here but in a different format than the way they were presented in the original text (1989, pp. 6-7).

20. Eichelberger (1989) cautions that differentiating the major positions of various philosophies or theoretical perspectives can be somewhat misleading. In practice, rarely does one pursue his/her work from a purely positivist, interpretive, etc. framework. There are many mixes, blends, and combination of activities reflected in contemporary research studies, each of which falls at various positions along the continuum of research perspectives but never rests fully at either end.

21. These long standing assumptions about ability grouping's perceived benefits have been usefully synthesized in a publication by the Massachusetts Board of Education (1990) and paraphrased here.

22. These generalizations have been culled by the present author from various reviews and single studies of the ability grouping literature.

23. From "Partnership in Parity: The Coalition for School Improvement," a 1989 Coalition for School Improvement publication, p. 1.

24. The principals of the two schools declining the initial invitation said that their school staff were either committed to other projects or that the study's purpose was not applicable - i.e. ability-based reading groups were not employed at the school or classroom level.

25. During this process, data analysis into categories and subcategories (Spradley, 1980), "data crunching" (Goetz and LeCompte, 1984), or data reduction (Miles and Huberman, 1984) procedures were begun. Some categories were immediately apparent, such as pre-assessment, formative, and summative criteria for student placement into within-class ability groups. Many of these categories describe what teachers say they do. Other categories emerged as themes and required a look for deeper meaning, as in the analysis of teachers' perceptions of the goals and expectations for student learning that they seek, the kind of information their evaluations of students yield, and what support factors are required in order to evaluate pupil learning in reading more effectively. These themes tend to be more interpretive.

26. It is interesting to note that the use of micro-computers as an enhancing and/or evaluative tool in the Reading classroom was rarely mentioned. Only two teachers, one in a basal and one using a novel-based approach reported information about computers. The teacher with the basal classroom was excited about the selected IBM software she was using to evaluate vocabulary development of her students. She also enthusiastically recounted how one student used the computer to compose a creative book report using graphics and word processing features.

27. Curriculum evaluation theory suggests that formative evaluation data can be used for summative decision-making purposes and that summative evaluation data can be cycled to support formative decision-making purposes.

28. Of course, cognition and affect are not mutually exclusive concepts. Many scholars have suggested that emotion and reason may not only go together but may, in fact, be stimulants to each other. This proposition has important implications for instructional theory and practice.

29. One impetus for continued scrutiny of the professional knowledge base of contemporary teacher education programs may be in the conserving, prescriptive types of knowledge that are perpetrated in much of the preservice curricula. Being critical of present ways in which public education is arranged, children are served, and teachers prepared, may prompt a response from one's colleagues or student peers akin to crossing a picket line at a large local employer.

30. An ad hoc group refers to a temporary group of students brought together for a particular purpose. When the purpose is accomplished, the group is usually disbanded. For example, teachers rather frequently report bringing students together who are having difficulty with a particular skill - e.g. cause and effect - to try to teach this principle to students in a more individualized or alternative manner.

BIBLIOGRAPHY

Ahmann J S and Glock M D (1981). Evaluating Student Progress. 6th edition. Boston: Allyn & Bacon.

Alpert J L (1974). Teacher behavior across ability groups: A consideration of the mediation of Pygmalion effects. Journal of Educational Psychology, 66, 348-353.

Alpert J L (1983). Do teachers adapt methods and materials to ability groups in reading? California Journal of Educational Research, Vol. XXVI, No. 3, 120-123.

Alexander J and Heathington B (1987). Assessing and Correcting Classroom Reading Problems. Glenview, IL: Scott, Foresman & Co.

Airasian P W (1984). Classroom assessment and educational improvment. Paper presented at the conference, Classroom Assessment: A Key to Educational Excellence, NWREL, Portland, OR.

Allington R (1983). The reading instruction provided readers of differing reading abilities. The Elementary School Journal, 83(5), 548-559.

Ames C (1984). Competitive, cooperative, and individualistic goal structures: A cognitive motivational analysis. R E Ames & C Ames (Eds.), Research on Motivation in Education, Vol I Student Motivation. New York: Academic Press.

Anderson L, Evertson C M and Brophy J E (1979). An experimental study of effective teaching in first grade reading groups. Elementary School Journal, 79, 193-223.

Anderson R C, Mason J and Shirey L (1984). The reading group: An experimental investigation of a labyrinth. Reading Research Quarterly, 20, 6-38.

Anderson R, Hiebert E, Scott J and Wilkinson I (1985). Becoming a Nation of Readers: The Report of the Commission on Reading. Washington, D.C.: National Institute of Education.

Apple M (Ed.) (1982). Culture and economic reproduction in education. London: Routledge & Kegan Paul.

Apple M (1982). Education and Power. London: Routledge & Kegan Paul.

Balow B and Curtin J (1966). Ability grouping of bright pupils. Elementary School Journal, 66, 321-326.

Bandura A (1977). Self-efficacy: Toward a unifying theory of behavioral change. Psychological Review, 84, 191-215.

Barnes S (1985). A study of classroom pupil evaluation: The missing link in teacher education. Journal of Teacher Education, 36(4), 46-49.

Barr R (1974). Instructional pace differences and their effect on reading acquisition, Reading Research Quarterly, 9, 526-554.

Barr R (1975). How children are taught to read: Grouping and pacing. School Review, 83, 479-498.

Bennett C (1979). Teaching students as they would be taught: The importance of cultural perspective. Educational Leadership, January, 259-268.

Bennett W (1986). First Lessons: A Report on Elementary Education in America, U.S. Department of Education. September.

Bereiter C and Scardamalia M (1984). Learning about Writing from Reading. Written Communication, 1 (2) April: 163-188.

Bettelheim and Zevan (1982). On Learning to Read: The Child's Fascination with Meaning. New York: Knopf.

Blakeslee T R (1983). The Right Brain: A New Understanding of the Unconscious Mind and Its Creative Powers. New York: Berkeley Books.

Block J (1971). Mastery Learning: Theory and Practice. New York: Holt, Rinehart & Winston.

Bloom B S (1956). Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain. New York: Longman.

Bloom B S (1976). Human Characteristics and School Learning. New York: McGraw-Hill.

Bloom B S (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher 13 (6), 4-16.

Bloom B S, Hastings H J and Madaus G F (1971). Handbook on Formative and Summative Evaluation of Student Learning. New York: McGraw-Hill.

Bloom B S, Madaus G and Hastings J (1981). Evaluation to Improve Learning. New York: McGraw-Hill.

Bloome D (1986). Building literacy and the classroom community. Theory Into Practice, Vol. XXV No. 2, Spring. Columbus, OH 71-76.

Bloome D (1989). The social construction of intertextuality on classroom literary learning. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Bloome D and Nieto S (1989). Children's understandings of basal readers. Theory Into Practice, Vol. XXVIII No. 4, Autumn. Columbus, OH 258-284.

Board of Education for the City of Etobicoke (1987). Making the Grade: Evaluating Student Progress. Scarborough, Ontario: Prentice-Hall Canada, Inc.

Bobbitt, F (1924). How to Make a Curriculum. Boston: Houghton Mifflin Co.

Bogdan R C and Biklen S K (1982). Qualitative Research for Education: An Introduction to Theory and Methods. Boston, MA: Allyn and Bacon, Inc.

Bolster A S (1983). Toward a more effective model of research on teaching. Harvard Educational Review, 53(3), 294-308.

Borg W R (1965). Ability grouping in the public schools: A field study. Journal of Experimental Education, 34(2), 1-97.

Borich G D and Madden S K (1977). Evaluating Classroom Instruction: A Sourcebook of Instruments. Reading, MA: Addison-Wesley.

Borko H (1978). An examination of some factors contributing to teachers' preinstructional classroom organization and management decisions. Paper presented at the annual meeting of the American Educational Research Association, Toronto.

Borko H and Eisenhart M (1989). Reading ability groups as literacy communities. Chapter 3 in David M. Bloome (Ed.) Classrooms and Literacy. Norwood, NJ: Ablex Publ.

Bowles S and Gintish (1976). Schooling in Capitalist America: Educational Reform and the Contradictions of Economic Life. New York: Basic Books.

Bradburn, Sudman, Norman B (1981). Improving Interview Method and Questionnaire Design. San Francisco: Jossey-Bass.

Brophy J (1982). Research on the self-fulfilling prophecy and teacher expectations. Paper presented at the Annual Meeting of the American Educational Research Association. (March)

Brophy J E & Good T (1970). Teachers' communication of differential expectations for children's classroom performance: Some behavioral data. Journal of Educational Psychology, 61, 365-374.

Brophy J E & Good T L (1986). Teacher behavior and student achievement. M C Wittrock, (Ed), Handbook of Research on Teaching, 3rd. edition, New York: Macmillan, 328-374.

Bruner J (1966). Toward a Theory of Instruction. Cambridge, MA: Belknap Press.

Burdsal C et al (1984). Behavioral dimensions in fourth, fifth, and sixth-grade students as rated by teachers. Journal of Clinical Psychology, 40, 172-185.

Calderhead J (1983). Teachers' Classroom Decision-Making. London: Holt, Rinehart & Winston.

Calderhead J (1983). A psychological approach to research on teachers' classroom decision-making. British Educational Research Journal, 7(1), 51-57.

Calfee R C and Drum P A (1986). Review of research on reading. M. Wittrock (Ed.), Handbook of Research on Teaching, Chapter 28, 3rd edition. New York: Macmillan.

Calkins L M (1986). The Art of Teaching Writing. Portsmouth, NH: Heinemann Educational Books.

Carbo M (1987). Deprogramming reading failure: Giving unequal learners an equal chance. Phi Delta Kappan, November, 197-202.

Carbo M (1987). Matching reading styles: Correcting ineffective instruction. Educational Leadership, October, 55-62

Carlson S B (1985). Creative classroom testing. Princeton, NJ: Educational Testing Service.

Carroll J (1963). A model of school learning. Teachers College Record, 64, 723-733.

Carson T (1990). What kind of knowing is critical action research? Theory Into Practice, Teacher as Researcher. Vol. XXIX, No. 3 Summer.

Cazden C (1985). Ability grouping and differences in reading instruction. In J Osborn, P Wilson and R Anderson (Eds.), Reading Education: Foundations for a Literate America. Lexington, MA: Lexington Books.

Chall J S (1983a). Learning to Read: The Great Debate. 2nd edition. New York: McGraw-Hill.

Charles C M (1983). Elementary Classroom Management, New York & London: Longman, Inc.

Charters W W (1924). Functional analysis as the basis for curriculum construction. Journal of Educational Research, 10, 214-221.

Clark C and Peterson P L (1986). Teacher thought processes. M. Wittrock (Ed.), Handbook of Research on Teaching, Chapter 20, 3rd edition. New York: Macmillan.

Coffman W E (1983). Testing in the schools: A historical perspective. Paper presented at the annual invitational conference of the Center for the Study of Evaluation, University of California, Los Angeles.

Cohen E (1986). Designing Groupwork: Strategies for the Heterogeneous Classroom. New York: Teachers College Press.

College Entrance Examination Boards (1985). Equality and Excellence: The Educational Status of Black Americans. New York: CEEB.

Comer J P (1988). Educating poor minority children. Scientific American, 259(5), 42-48.

Connelly F M and Clandinin D J (1985). Personal practical knowledge and the modes of knowing: Relevance for teaching and learning. NSSE Yearbook, 84(2), 174-198.

Connelly F M and Elbaz F (1980). Conceptual bases for curriculum thought: A teachers' perspective. In A W Foshay (Ed.). Considered Action for Curriculum Improvement, Alexandria, VA: ASCD, 95-119

Cornell, E (1936). Effects of ability grouping determinable from published studies. In G M Whipple (Ed.), Adapting the School to Individual Differences. National Social Studies Yearbook, Bloomington, IL: Public School Publishing Co.

Crooks T (1988). The impact of classroom evaluation practices on students. Review of Educational Research, Vol. 58, No. 4, Winter: University of Otago, 438-481.

Cuban L (1984). How Teachers Taught: Constancy and Change in American Classrooms 1890-1980. New York: Longman.

Dar Y (1985). Teachers' attitudes toward ability grouping: Educational considerations and social organizational influences. Interchange, 16(2), 17-38.

Dar Y and Resh N (1986). Classroom intellectual composition and academic achievement. American Educational Research Journal, 23(3).

Denzin, N K (1978). The logic of naturalistic inquiry. In N K Denzin (Ed.). Sociological Methods: A Sourcebook. New York: McGraw-Hill.

Dewey J (1910). How We Think. Boston: Heath.

Dewey J (1913). Interest and Effort in Education. Copyright 1975 by Center of Dewey Studies, Southern Illinois: University at Carbondale.

Dewey J (1934). Art as Experience. New York: Minton, Balch and Company.

234

Dorr-Bremme D W, Herman J. (1986). Assessing school achievement: A profile of classroom practices. Los Angeles: Center for the Study of Evaluation, UCLA Graduate School of Education.

Doyle W (1979). Making managerial decisions in classrooms. In D L Duke (Ed.), Classroom Management. The Seventy-Eighth Yearbook of the National Society for the Study of Education, Part II. Chicago: University of Chicago Press.

Doyle W (1986). Classroom organization and management. Wittrock, M C (Ed.). Handbook of Research on Teaching, 3rd edition, New York: Macmillan.

Dreeban R (1968). On What is Learned in School. Reading, MA: Addison-Wesley.

Dreeban R (1984). First-grade reading groups: Their formation and change. In P Peterson, L Wilkinson and M Hallinan (Eds.), The Social Context of Instruction: Group Organization and Group Processes. Orlando, FL: Academic Press.

Dreeban R and Barr R (1988). The formation and instruction of ability groups. American Journal of Education. November, Chicago: The University of Chicago.

Driscoll, M P (1986). The relationship between grading standards and achievement: A new perspective. Journal of Research and Development in Education, 19(3), 13-17.

Duffy G G (1981). Teacher effectiveness research: Implications for the reading profession. In M L Kamil (Ed.), Directions in Reading: Research and Instruction. Washington, D.C.: National Reading Conference. 113-136.

Dunkin M J (Ed.) (1987). The International Encyclopedia of Teaching and Teacher Education. New York: Pergamon Press.

Durkin D (1982). A study of poor black children who are successful readers. Reading Education Report No. 33. Urbana: University of Illinois, Center for the Study of Reading.

Dweck C S and Elliott E S (1983). Children's theories of intelligence: Consequences for learning. Learning and Motivation in the Classroom. Hillsdale, New Jersey: Erlbaum.

Eash J and Snow R (1971). Pygmalion Reconsidered. Worthington, OH: Chas. A.

Ebel R L (1980). Evaluation of Students: Implications for Effective Teaching, Educational Evaluation and Policy Analysis, 2(1), Jan-Feb 47-51.

Eder D (1981). Ability grouping and a self-fulfilling prophecy: Micro analysis of teacher-student interaction. Sociology of Education, 54, 151-162.

Eder D (1982). Differences in communicative styles across ability groups. Board of Regents of the University of Wisconsin.

Edwards, A D and Mercer, N (1987). Common Knowledge, New York: Methuen.

Eickelberger T (1989). Disciplined inquiry: Understanding and doing social research. Philosophical Foundations of Knowledge, Chapter 1. New York: Longman, Inc., 1-20.

Einhorn H J and Hogarth R M (1978). Confidence in judgment: Persistence of the illusion of validity. Psychological Review, 86, 465-485.

Eisner E W (1975). Educational connoisseurship and educational criticism: Their form and function in educational evaluation. Also, The perceptive eye: Toward the reform of educational evaluation. Invited address at the annual meeting of the AERA, Washington, DC. March

Eisner E W (1985). The Educational Imagination: On the Design and Evaluation of School Programs. New York: Macmillan Publishing Company.

Emmer E, Evertson C and Anderson L (1980). Effective classroom management at the beginning of the school year. Elementary School Journal, 80, 219-31.

Erickson F (1986). Qualitative methods in research on teaching. M. Wittrock (Ed.), Handbook of Research on Teaching, Chapter 5, 3rd edition. New York: Macmillan.

Esposito D (1973). Homogeneous and heterogeneous ability grouping: Principal findings and implications for evaluating and designing more effective educational environments. Review of Educational Research, 43, 163-179.

Evertson C M, Hawley W D and Zlotnik M (1985). Making a difference in educational quality through teacher education. Journal of Teacher Education, May-June, 2-12.

Fair J W et al (1980). Teacher interaction and observation practices in the evaluation of student achievement. Toronto: Ontario Ministry of Education.

Felmlee D and Eder D (1983). Contextual effects in the classroom: The impact of ability group on student attention. Sociology of Education, 56, 77-87.

Fenstermacher, G D (1987). On understanding the connections between classroom research and teacher change. Theory Into Practice, Vol. XXVI, No. 1, Winter, Columbus, OH, 3-7.

Fleming M and Chambers B (1983). Teacher-made tests: Windows to the classroom. In W E Hathaway (Ed.), Testing in the Schools: New Directions for Testing and Measurement, 19, San Francisco: Jossey-Bass. 29-38.

Florio S and Walsh M (1980). The teacher as colleague in classroom research. In H. Trueba, G Guthrie and K Au (Eds.) Culture in the bilingual classroom: Studies in classroom ethnography. Rawley, MA: Newbury House.

Floyd C (1954). Meeting children's reading needs in the middle grades: A preliminary report. Elementary School Journal, 55 99-103.

Fraatz J M (1987). The Politics of Reading: Power, Opportunity, and Prospects for Change in America's Public Schools. New York: Teachers College Press.

Gage N L (1985). Hard gains in the soft sciences: The case of pedagogy. Bloomington, IN: Phi Delta Kappa, Center on Evaluation, Development and Research.

Gamoran A (1984). Egalitarian versus elitist use of ability grouping. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Gardner H (1983). Frames of Mind: A Theory of Multiple Intelligences. New York: Basic Books.

Gardner J (1962). Excellence: Can We be Equal and Excellent, Too? New York: Harper.

Ghory W J (1977). Alternative educational environments: Marginal learner perceptions of curriculum conditions in public alternative high schools. Unpublished doctoral dissertation. University of Massachusetts, Amherst.

Gil D (1980). The decision-making and diagnostic processes of classroom teachers. (Research Series No. 71). East Lansing, MI: Institute for Research on Teaching, Michigan State University.

Giroux H (1983). Theory and Resistance in Education: A Pedagogy for the Opposition. London: Heinemann.

Glaser R (1963). Instructional technology and the measurement of learning outcomes: Some questions. American Psychologist, 18, 519-521.

Glatthorn A (1987). Curriculum Leadership. Glenview, IL: Scott, Foresman and Company.

Good T and Brophy J (1987). Looking in Classrooms. 4th edition. New York: Harper & Row.

Goodlad J (1983). A Place Called School: Prospects for the Future. New York: McGraw Hill.

Goodlad J I (1966). The development of a conceptual system for dealing with problems of curriculum and instruction. Washington, D.C.: Report sponsored by the Cooperative Research Program of the Office of Education, United States Department of HEW.

Goodlad J I and Oakes J (1988). We must offer equal access to knowledge. Educational Leadership, February, 16-22.

Goodlad J I, Sirotnik K A and Overman B C (1979). An overview of a study of schooling. Phi Delta Kappan, 61, 174-178.

Goodman K S (1989). Access to literacy: Basals and other barriers. Theory Into Practice, Vol. XXVIII, No. 4, Autumn: Columbus, OH, 300-306.

Grant C (1979). Classroom socialization: The other side of a two-way street. Educational Leadership, April, 471-473.

Grant G and Briggs J (1983). Today's children are different. Educational Leadership, March, (4), 6-9.

Grant L and Rothenberg J (1986). The social enhancement of ability differences: Teacher-student interactions in first and second grade reading groups. The Elementary School Journal, 87 (1) Chicago: The University of Chicago, 29-49.

Green-J L (1983). Research on teaching as a linguistic process: A state of the art. In E W Gordon (Ed.), Review of Research in Education, Washington, DC: American Educational Research Association, 10, 152-252.

Grumet M R (1988). Bittermilk: Women and Teaching. Amhest: University of Massachusetts.

Grumet M R (1989). Generations: Reconceptualist curriculum theory and teacher education. Journal of Teacher Education. Vol. XXXX, No. 1, Jan-Feb.

Guba E and Lincoln Y (1981). Effective Evaluation. San Francisco: Jossey-Bass.

Gullickson A R (1985). Student evaluation techniques and their relationship to grade and curriculum. Journal of Educational Research, 79, 96-100.

Gullickson A R and Ellwein M C (1985). Post hoc analysis of teacher made tests: The goodness of fit between prescription and practice. Educational Measurement: Issues and Practice, 4(1), 15-18.

Gumperz J J (1986). Interactional sociolinguistics on the study of schooling. In J Cook-Gumperz (Ed.), The Social Construction of Literacy, New York: Cambridge University Press, 45-68.

Gwaltney J L (1980). Drylongso: A Self-Portrait of Black American. New York: Random House.

Haberman M (1984). An evaluation of the rationale for required teacher education: Beginning teachers with and without teacher preparation. Paper presented for the National Commission on Excellence in Teacher Education.

Habermas J (1972). Knowledge and Human Interests (J Shapiro, Trans.) Boston: Beacon Press.

Hale J E (1982). Black Children: Their Roots, Culture, and Learning Styles. Provo, Utah: Brigham Young University Press.

Halkes R and Olson J K (Eds.) 1984. Teacher thinking: A new perspective on persisting problems in education (123-133). Lisse: Swets and Zeitlinger B.V.

Haller E J (1985). Pupil race and elementary school ability grouping: Are teachers biased against black children? American Educational Research Journal, 22, 485-499.

Hallinan M T and Sorenson A B (1983). The formation and stability of instructional groups. American Sociological Review, 48, 838-851.

Hambleton R R (1982). Advances in criterion-referenced testing technology. D C Reynolds and T Gutkin (Eds.), Handbook of School Psychology, New York: Wiley.

Harlan W (Ed.) (1978). Evaluation and the Teacher's Role. London: Macmillan.

Harris L A and Smith C B (1986). Reading Instruction: Diagnostic Teaching in the Classroom. New York: Macmillan Publishing Company.

Harrison E A (1989). Ability grouping: Practices and perceptions of elementary school teachers. A dissertation presented to the Graduate School of University of Massachusetts, Amherst.

Hart L (1986). New concepts in testing gaining ground, may replace antique ways. Cutting Edge, 9, 1.

Heath S B (1983). Ways with Words, Language, Life, and Work in Communication and Classrooms. Cambridge, England: Cambridge University Press.

Heathers G (1969). Grouping. In R Ebel (Ed.), Encyclopedia of Educational Research, 4th edition. New York: Macmillan. 559-570.

Hermann J and Dorr-Bremme D W (1982, March). Assessing students: Teachers' routine practices and reasoning. Paper presented at the annual meeting of the American Educational Research Association.

Hesse M (1980). Revolutions and Reconstructions in the Philosophy of Science. Bloomington: Indiana University Press.

Hiebert E (1983). An examination of ability grouping in reading instruction. Reading Research Quarterly, 18, 231-253.

Hillson M, Jones J C, Moore J W and Van Devender F (1964). A controlled experiment evaluating the effects of a non-graded organization on pupil achievement. Journal of Educational Research, 57, 548-550.

Hirst P (1980). Knowledge and the Curriculum. London: Routledge & Kegan Paul, 1980.

Howe II, H and Edleman M W (1985). Barriers of excellence: Our children at risk. Boston: National Coalition of Advocates for Students, 93.

Hopkins K and Stanley J (1981). Educational and Psychological Measurement and Evaluation. Englewood Cliffs, NJ: Prentice-Hall.

House E (1978). Assumptions underlying evaluation models. Educational Researcher, March, 4-12.

Hutchinson T and Fortune J (1982). Decision-Maker Oriented Evaluation Methodology. Amherst, MA: Center for Research and Evaluation Methods.

Illich I (1972). Deschooling Society. New York: Harper and Row.

Ittleson, W H and Cantril H (1954). Perception: A Transactional Approach. New York Doubleday Papers in Psychology.

Jackson P W (1968). Life in Classrooms. New York: Holt, Rinehart & Winston.

Jewell M G and Zintz M V (1986). Learning to Read Naturally. Dubuque, Iowa: Kendall/Hunt Publishing.

Johnston P H (1983). Reading comprehension assessment: A cognitive basis. Newark, DE: International Reading Association.

Johnston P H and Winograd P N (1985). Passive failure in reading. Journal of Reading Behavior, 17, 279-301.

Joint Committee on Standards for Educational Evaluation (1981). Standard for Evaluation of Educational Programs, Projects, and Materials. New York: McGraw-Hill.

Jones D (1977). Curriculum Targets in the Elementary School. Englewood Cliffs, NJ: Prentice-Hall.

Kant E (1952). The critique of pure reason. In R M Hutchins (Ed.) The Great Books of the Western World, 42, Chicago: Encyclopedia Britannica. (Original work published 1787.)

Katz M B (1975). Class, Bureaucracy and Schools: the Illusion of Educational Change in America. New York: Praeger Publishers.

Kellagan T, Madaus G F and Airasian P W (1982). The Effects of Standardized Testing. Boston: Khuwer-Nijhoff.

Kelley H H and Michela J L (1980). Attribution theory and research. Annual Review of Psychology, 31, 457-501.

Kirk J and Miller M L (1986). Reliability and Validity in Qualitative Research. Beverly Hills: Sage.

Kirst M (1984). Who Controls Our Schools: American Values in Conflict. Stanford CA: Stanford Alumni Association.

Klein M F (1989). Curriculum Reform in the Elementary School: Creating Your Own Agenda. New York: Teachers College Press.

Komarowsky M (1946). Cultural contradictions and sex roles. American Journal of Sociology, 52, 184-189.

Kounin J (1970). Discipline and Group Management in Classrooms. New York: Holt, Rinehart & Winston.

Kourilsky M (1974). An adversary model for educational evaluation. Evaluation Comment, 4.

Kuhn T S (1962). The Structure of Scientific Revolutions. Chicago: University of Chicago Press.

Kulik C C and Kulik J A (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. American Educational Research Journal, 19, 415-428.

Kulik C C and Kulik J A (1984, August). Effects of ability grouping on elementary school pupils: A meta-analysis. Paper presented at the annual meeting of the American Psychological Association. Toronto.

LaPoint A (1986). The state of instruction in reading and writing in U.S. elementary schools. Phi Delta Kappan. October, 135-138.

Lazar-Morris, Polin L, May R and Barry L (1980). A review of the literature on test use. Los Angeles: UC Center for the Study of Evaluation, ERIC No. ED 204 411.

Lemlech J (1988). Classroom Management: Methods and Techniques for Elementary and Secondary Teachers. 2nd edition. White Plains, NY: Longman, Inc.

Levine M (1974). Scientific method and the adversary. American Psychologist, September, 666-677.

Lincoln Y and Guba E (1985). Naturalistic Inquiry, Beverly Hills: Sage.

Lieberman A and Miller L (1984). Teachers, their world and their work: Implications for school improvement. Alexandria, VA: Association for Supervision and Curriculum Development.

Locke L F (1984). Research on teaching teachers: Where we are now. Journal of Teaching in Physical Education, 2(2), 1-86.

Maehr M L (1976). Continuing motivation: An analysis of a seldom considered educational outcome. Review of Educational Research, 46, 443-462.

Malinowski B (1960). A Scientific Theory of Culture and Other Essays, New York: Oxford University Press.

Manen van M (1990). Beyond assumptions: Shifting the limits of action research. Theory Into Practice, Vol. XXIX, No. 3 Summer: Columbus, OH, 152-157.

Marland P W (1979). A study of teachers' interactive thoughts. Doctoral dissertation, University of Alberta.

Marshal J P (1983). Teaching's Most Taxing Traditions: Reflections on Evaluation and Grading. University of Texas at Austin. Occasional Paper. Spring

Mason J and Osborn J (1982). When do children begin "reading to learn"?: A survey of classroom reading instruction practices in grades two through five. (Technical Report No. 261). Urbana: University of Illinois, Center for the Study of Reading.

Massachusetts Board of Education (1990). Structuring schools for student success: A focus on ability grouping. Publication #16, January, 1-23.

McClendon J (1986). Developing writers in an itermediate-grade classroom. Theory Into Practice, Vol. XXV, No. 2, Spring: Columbus, OH, 117-123.

McCutcheon G and Jung B (1990). Alternative perspectives on action research. Theory Into Practice, Teacher as Researcher, Vol. XXIX, No. 3, Summer.

McNair K (1978). Capturing inflight decisions: Thoughts while teaching. Educational Research Quarterly, 3, 26-42.

Mead M (1951). The School in American Culture, Cambridge, MA: Harvard University Press.

Meehan E (1969). Value Judgment and Social Science. Homewood, IL: Dorsey Press.

Mehrens W and Lehmann I J (1984). Measurement and Evaluation in Education and Psychology. New York: CBS College Publishing.

Meier K J, Stewart J Jr. and England R E (1989). Race, Class and Education: The Politics of Second-Generation Discrimination. Madison, Wisconsin: University of Wisconsin Press.

Michaels S and Cook-Gunperz T (1979). A study of sharing time with first grade students: Discourse narratives in the classroom. Proceedings of the Fifth Annual Meeting. Berkeley, CA: Berkeley Linguistics Society, 647-660.

Miles M B and Huberman A M (1984). Drawing valid meaning from qualitative data: Towards a shared craft. Educational Researcher, 13(5), 20-30.

Mill J S (1906). A System of Logic. London: Longman, Inc. (Original work published 1843).

Morinne-Dershimer G (1978-9). How teachers "see" their pupils. Educational Research Quarterly, 3, 83-99.

Morinne-Dershimer G and Vallance E (1975). A study of teacher and pupil perceptions of classroom interactions. Beginning Teacher Evaluation Study. San Francisco: Far West Laboratory, Technical Report 75-11-6.

Morris V P (1969). An evaluation of pupil achievement in a nongraded primary plan after three, and also five years of instruction. Dissertation Abstracts, 29, 3809-A, University Microfilms No. 69-7352.

Moses P J (1966). A study of inter-class ability grouping on achievement in reading. Dissertation Abstracts, 26, 4342.

Murnane R J and Phillips B R (1978). Effective teachers of inner-city children: Who they are and what they do. Princeton, NJ: Mathematica Policy Research.

National Assessment of Educational Progress (1985). The Reading Report Card: Progress Toward Excellence in Our Schools; Trends in Reading Over Four National Assessments, 1971-1984. Princeton, NJ: Educational Testing Service.

National Education Association (1968). Ability grouping research summary. Washington, D.C.: NEA.

National Education Association (1982). Excellence in Our Schools, Teacher Education: An Action Plan. Washington DC: NEA.

National Education Association of the United States (1895). Journal of Proceedings and Addresses; Session of the Year 1895, St. Paul, MN.

National Institute of Education (1975). Teaching as clinical information processing. (Report of Panel 6, National Conference on Studies in Teaching). Washington DC: NIE.

Natriello G (1987). The impact of evaluation processes on students. Educational Psychologist, 22, 155-175.

Newcomer P L (1975). Accuracy of teacher judgments. Elementary English, 52, 704-707.

Newport J F (1969). The Joplin Plan: The score. Reading Teacher, 21, 158-162.

Nieto, Sonia (forthcoming). Affirming Diversity: The Sociopolitical Context of Multicultural Education. New York: Longman, Inc.

Nisbett R and Ross L (1980). Human influence: Strategies and shortcomings of social judgments. Englewood Cliffs, NJ: Prentice-Hall.

Nitko A (1984). Defining criterion-referenced test. In Ronald A Berk (Ed.), A Guide to Criterion-Referenced Test Construction. Baltimore: Johns Hopkins University Press, 8-28.

Nobles W W (1974). Africanity: Its role in black families. Black Scholar, 5, 10-17.

Oakes J (1985). Keeping Track: How Schools Structure Inequality. New Haven: Yale University Press.

Otto H and Miller W (1930). Analysis of experimental studies in homogeneous grouping. Journal of Educational Research, 95-102.

Owens T (1973). Education evaluation by adversary proceedings. In E. House (Ed.), School Evaluation: The Politics and Process. Berkeley, CA: McCutchan.

Owings R A et al (1980). Spontaneous monitoring and regulation of learning: A comparison of successful and less successful fifth graders. Journal of Educational Psychology, 72, 250-256.

Parlett M and Hamilton D (1977). Evaluation as illumination: A new approach to the study of innovatory programs. In G V Glass (Ed.), Evaluation Studies Review Annual (1), Beverly Hills, CA: Sage.

Patton, M (1980). Qualitative Evaluation Methods. Beverly Hills, CA Sage.

Pearson P (Ed.) (1984). Handbook of Reading Research. New York: Longman, Inc.

Percell C (1977). Education and Inequality: A Theoretical and Empirical Synthesis. New York: The Free Press.

Pinar W F (Ed.) (1978). Curriculum Theorizing: The Reconceptualists. Berkeley CA: McCutchan.

Popham W (1975). Educational Evaluation. Englewood Cliffs, NJ: Prentice-Hall.

Popham W (1978). Criterion-Referenced Measurement. Englewood Cliffs, NJ: Prentice-Hall.

Prendergast P (1983). Institutional Self-Study Report. North Adams State College, North Adams, MA

Quist D et al (1985). Interstate Certification Commission Exit Report on Five-Year Accreditation of Teacher Education Programs, North Adams State College, North Adams, MA. Spring.

Puro P and Bloome D (1987). Understanding classroom communication. Theory Into Practice. Vol. XXVI, No. 1, Winter: Columbus, OH, 26-31.

Reese W L (1980). Dictionary of Philosophy and Religion. Atlantic Highlands, NJ: Humanities.

Resnick D P and Resnick L B (1977). The nature of literacy: An historical exploration. Harvard Educational Review, 47, 370-385.

Riessman F (1976). Students' learning styles: How to determine, strengthen, and capitalize on them. Chapter 5, in The Inner-City Child by Frank Riessman. New York: Harper & Row.

Rinne C (1984). Attention: The Fundamentals of Classroom Control. Columbus, OH: Charles E. Merrill.

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

Rosenholtz S J and Simpson C (1984). Classroom organization and student stratification. Elementary School Journal, 85, 21-37.

Rosenshine B and Stevens R (1984). Classroom instruction in reading. In P D Pearson (Ed.), Handbook of Reading Research. New York: Longman, 745-798.

Rosenshine B and Stevens R (1986). Teaching functions. Handbook of Research on Teaching, 3rd Ed., New York: Macmillan, 376-391.

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

Rowan S and Miracle A (1983). Systems of ability grouping and the stratification of achievement in elementary schools. Sociology of Education, 56, 133-144.

Rugg H et al (1927). The foundations of curriculum-making. Chapter 1 in The Foundation of Curriculum-Making. Twenty-sixth Yearbook of the National Society for the Study of Education, Part II. Bloomington, IL.

Russo N A (1978). Capturing teachers' decision policies: An investigation of strategies for teaching reading and mathematics. Paper presented at the annual meeting of the American Educational Research Association. Toronto.

Salmon-Cox L (1981). Teachers and standardized achievement tests: What's really happening? Phil Delta Kappa, 62, 631-634.

Sax G (1980). Principles of Educational and Psychological Measurement and Evaluation. Belmont CA: Wadsworth.

Saylor J, Alexander W and Lewis A (1981). Curriculum Planning for Better Teaching and Learning, 4th edition. New York: Holt, Rinehart & Winston.

Schon D (1983). The Reflective Practitioner: How Professionals Think in Action. New York: Basic Books.

Schubert W (1987). Curriculum: Perspective, Paradigm, and Possibility. New York: Macmillan.

Schunk D (1985). Self efficacy and classroom learning. Psychology in Schools, 22, 208-223.

Schwartz F (1981). Supporting or subverting learning: Peer group patterns in four tracked schools. Anthropology and Education Quarterly, 12(2).

Scriven M (1967). The methodology of evaluation. AERA Monograph Series on Curriculum Evaluation, Book I, Robert E Stake (Ed.) Chicago: Rand McNally & Co.

Scriven M (1973). Goal-free evaluation. In School Evaluation: The Politics and Process. E. House (Ed.). Berkeley, CA: McCutchan.

Shavelson R J (1976). Teachers' decision-making. In N L Gage (Ed.), The Psychology of Teaching Methods, Yearbook of the National Society for the Study of Education. Chicago: University of Chicago Press, 372-414.

Shavelson R J (1983). Review of research on teachers' pedagogical judgments, plans, and decisions. Elementary School Journal, 83(4), 392-413.

Shavelson R J and Stern P (1981). Research on teachers' pedagogical thoughts, judgments, decisions, and behavior. Review of Educational Research, 51, 455-498.

Shulman L S (1980). Test design: A view from practice. In E L Baker and E S Quellmalz (Eds.), Educational Testing and Evaluation. Los Angeles, CA: Sage, 63-73.

Shulman L S and Elstein A S (1975). Studies of problem solving, judgment, and decision making: Implications for educational research. In F N Kerlinger (Ed.), Review of Research in Education, Vol. 3.

Simon H A (1957). Models of Man: Social and Rational: Mathematical Essays. New York: Wiley.

Sinclair R J and Ghory W J (1979). Curriculum as Environments for Learning: A Practical Meaning and Model. Paper presented at the annual meeting of the American Educational Research Association in San Francisco, CA. (April)

Sinclair R J and Ghory W J (1987). Reaching Marginal Students: A Primary Concern for School Renewal. Chicago: McCutchan.

Sjoberg G (1975). Politics, ethics, and evaluation research. In M Guttentag and E L Struening (Eds.), Handbook of Evaluation Research (2), Beverly Hills, CA: Sage.

Slavin R E (1987). Ability grouping and student achievement in elementary schools: A best-evidence synthesis. Review of Educational Research, 57, 293-337.

Smith F (1985). Reading Without Nonsense. 2nd edition. New York: Teachers College Press.

Smith P L (1982). The Problem of Values in Educational Thought. Ames: Iowa State University Press.

Smith R and Tyler R (1942). Appraising and Recording Student Progress. New York: Harper & Row.

Spindler G (1982). Doing the Ethnography of Schooling. New York: Holt, Rinehart & Winston.

Spradley J P (1979). The Ethnographic Interview. New York: Holt, Rinehart & Winston.

Spradley J P (1980). Participant Observation. New York: Holt, Rinehart & Winston.

Sroufe G. (1977). Evaluation and politics. In J D Scriber (Ed.), The Politics of Education: Seventy-sixth yearbook of the National Society for the Study of Education, Part II. Chicago: University of Chicago Press.

Stake R (1967). The countenance of educational evaluation. Teachers College Record, 68, 523-540.

Stallings J and Hentzell (1978). What do we know about teaching and learning in urban schools? Paper prepared for the Urban Education Program, CEMREL, Inc.'s National Conference on Urban Education. St. Louis, MO, July.

Stallings J A (1987). Are we evaluating what we value? Action in Teacher Education, The Journal of the Association of Teacher Educators, 9(3), 1-3.

Stern P and Shavelson R (1981, June). The relationship between teachers' grouping decisions and instructional behaviors: An ethnographic study of reading instruction. Resources in Education. (ERIC Document Reproduction Service No. ED 201 971).

Sternberg R J (1986). Intelligence Applied. San Diego, CA: Harcourt Brace.

Sternberg R J (1986). Appraising and Recording Student Progress. New York: Harper & Row.

Stiggins R J (1988). Revitalizing Classroom Assessment: The Highest Instructional Priority. In Phi Delta Kappan. (January): 363-368.

Stiggins R J (1989). Teacher training in assessment: overcoming the neglect. Buros Nebraska Symposim in Measurement and Testing, Vol. 7, December, New York: Earl Baum, 1-21.

Stiggins R J and Bridgeford N J (1985). The ecology of classroom assessment. Journal of Educational Measurement, 271-286.

Stiggins R J, Conklin N F and Bridgeford et al (1986). Classroom assessment: A key to educational excellence. In Educational Measurement: Issues and Practice, 5, 5-17.

Storey A (1970). The measurement of classroom learning. Chicago: Science Research Associates.

Stufflebeam D et al (1971). Educational Evaluation and Decision-Making. Itasca, IL: F E Peakcock.

Swing S R and Peterson P L (1982). The relationship of student ability and small-group instruction to student achievement. American Educational Research Journal, 19, 259-274.

Taba H (1962). Curriculum Development, Theory and Practice. New York: Harcourt, Brace and World, Inc.

Talmadge H (1982). Evaluation of programs. In H. Mitzel (Ed.). Encyclopedia of Educational Research. 5th edition, Vol 2. New York: The Free Press, 592-608.

Taylor D (1988). Ethnographic educational evaluation for children, families, and schools. Theory into Practice, Winter XXVII (1), 67-78.

Taylor P H (1970). How teachers plan their courses. Slouth: National Foundation for Educational Research in England and Wales.

Taylor S J and Bogdan R (1984). Introduction to Qualitative Research Methods. New York: John Wiley & Son.

Theissen D and Moorhead R (1985). More than marks: What teachers say about student evaluation. Ontario Public School Teachers' Federation. 1260 Bay Street, Toronto, Ontario. November.

Thorndike E L (1906). The Principles of Teaching. New York: A G Seiler.

Thorton J and Wright J (1963). Secondary School Curriculum. Columbus, OH: Charles E. Merrill, INc.

Trimble K (1988). On the Wrong Track: Equal Access to Knowledge and Ability Grouping in Secondary Social Studies Classes. Doctoral Dissertation, University of Massachusetts, Amherst.

Trow M (1970). Participant observation and interviewing: A comparison. In W J Filstead (Ed.), Qualitative Methodology. Chicago: Markham.

Tunnell M O and Jacobs J S (1989). Using "real" books: Research findings on literature based reading instruction. The Reading Teacher, March, 470-477.

Tversky A and Kahneman D (1974). Judgment under uncertainty: Heuristics and biases. Science 185: 1124-31.

Tyack D (1974). The One Best System: A History of American Urban Education. Cambridge: Harvard University Press.

Tyler R W (1949). Basic Principles of Curriculum and Instruction: Syllabus for Education 360. Chicago: University of Chicago Press.

Tyler R W (1987). Education reforms. In Phi Delta Kappan. (December): 277-280.

Tyler R W (1988). What we know about learning and how that influences curriculum renewal. In Renewing School Curriculum: Concerns for Equal and Quality Education, Robert L. Sinclair and Sonia M. Nieto (Eds.) Coalition for School Improvement, University of Massachusetts, Amherst.

Tyler R W and Wolf R M (Eds.) (1974). Crucial Issues in Testing. Berkeley, CA: McCutchan.

Varner G F (1922). Can teachers select bright and dull pupils? Journal of Educational Research, 6, 126-131.

Vellutino, F R and Scanlon D M (1986). Experimental evidence for the effects of instructional bias on word identification. Exceptional Children, Oct. 145-155.

Vinsonhaler J F (1979). The consistency of reading diagnosis. East Lansing, MI: Institute for Research on Teaching, Michigan State University. Research Series No. 28 (June).

Wahlstrom M, Regan E and Jones L (1982). Teacher beliefs and the assessment of student achievement. In K Leithwood (Ed.), Studies in Curriculum Decision-Making. Toronto: OISE Press, 27-34.

Walberg H J (1990). Productive teaching and instruction: Assessing the knowledge base. In Phi Delta Kappan (February), 470-478.

Webb N M (1982). Student interaction and learning in small groups. Review of Educational Research, 52, 421-445.

Weinshank A B (1980). Investigations of the diagnostic reliability of reading specialists, learning disabilities specialists, and classroom teachers: Results and implications. (Research Series No. 72) East Lansing: Michigan State University, Institute for Research on Teaching, ERIC No. ED 189574.

Weinstein R (1976). Reading group membership in first grade: Teacher behaviors and pupil experience over time. Journal of Educational Psychology, 68, 103-116.

Weiss Carol H (1975). Evaluation research in the political context. In M. Guttentag and E L Streuning (Eds.), Handbook of Evaluation Research (2) 13-26, Beverly Hills, CA: Sage.

Whipple G M (Ed.) (1935). Educational Diagnosis. The Thirty-Fourth Yearbook of the National Society for the Study of Education. Bloomington, IL: Public School Publishing Company.

Whipple G M (Ed.) (1937). The Teaching of Reading: A Second Report. The Thirty-Sixth Yearbook of the National Society for the Study of Education, Part I. Bloomington, IL: Public School Publishing Company.

Whitehead A N (1929). The Aims of Education. New York: Macmillan.

Whitmer S P (1983). A descriptive multimethod study of teacher judgment during the marking process. (Research Series No. 122) East Lansing, MI: Michigan State University, Institute for Research on Teaching, ERIC No. ED234052.

Williams L V (1983). Teaching for the Two-Sided Mind. Englewood Cliffs, NJ: Prentice-Hall.

Williamson P M (1983). The Meaning of "A Curriculum" in Contemporary Practice. Doctoral Dissertation, University of Massachusetts, Amherst.

Wolf R (1975). Trial by jury: A new evaluation method. Phil Delta Kappan, 57.

Wood R and Napthali W (1975). Assessment in the classroom: What do teachers look for? Educational Studies, 1(3), 151-161.

Worthen B R and Sanders J R (1973). Educational Evaluation: Theory and Practice. Worthington, OH: Chas A Jones.

Worthen B R and Sanders J R (1987). Educational Evaluation: Alternative Approaches and Practical Guidelines. New York: Longman.

Wright C and Nuthall F (1970). The relationships between teacher behaviors and pupil achievement in three experimental elementary science lessons. American Educational Research Journal, 7, 477-492.

Yinger R J (1977). A study of teacher planning: description and theory development using ethnographic and information processing methods. Doctoral dissertation, Michigan State University.

Yinger R (1980). A study of teacher planning. Elementary School Journal, 80, 107-127.

Zahorik J A (1968). Classroom feedback behavior of teachers. Journal of Educational Research, 62, 147-150.

Zahorik J A (1975). Teachers' planning models. Educational Leadership, 33, 133-139.