



RETURNING MATERIALS:
Place in book drop to
remove this checkout from
your record. FINES will
be charged if book is
returned after the date
stamped below.

--	--	--

A DESCRIPTIVE MULTIMETHOD STUDY
OF TEACHER JUDGMENT DURING THE
MARKING PROCESS

By
Sylvia Pratt Whitmer

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Elementary and Special Education

1981

ABSTRACT

A DESCRIPTIVE MULTIMETHOD STUDY OF TEACHER JUDGMENT DURING THE MARKING PROCESS

By

Sylvia Pratt Whitmer

Persistent dissatisfaction with traditional teacher marks, A B C D E, prompted this study. The purpose was to generate a description of the judgment processes which allow a teacher to distill a single symbol, a mark, from the diverse activities of the classroom to represent pupil progress. The subjects were five experienced, upper elementary teachers from a typically achieving district in Michigan. The framework for the investigation came from the field of human judgment. Four established methods were applied: process tracing, policy capturing, attribution theory and utility theory. Field data included taped interviews, record books, marks and predicted marks across a year. Data were organized into a composite case and five teacher cases. Quantitative and qualitative analysis included multiple regression, Pearson and partial correlations, frequency counts and categorization and coding of verbatim interview protocols.

The findings revealed that teachers generally use procedural and contingency rules to guide a three-stage process of collecting data, weighting and assigning it to preordained categories, and choosing between categories when cumulated scores fall in zones of uncertainty or failure. The teachers' primary tool of inference at the procedural level is the record book, and the dominant judgment factor is task completion. The completion factor has a variable weight in the judgment process dependent upon task difficulty and student ability. Judgment factors which impact teacher choice in zones of uncertainty include

ability, effort, task difficulty, home support and classroom behavior/physical maturity. Of these, effort has the most weight. The contingency focus is upon those factors which promote individual and group task completion; the primary contingency tools of inference are checks, minuses and pluses. The marking process is restricted to classroom functions, a conclusion suggesting that previous marking studies have made inappropriate assumptions about teacher judgment processes. Formative marks serve a feedback function, but summative marks do not. A marking judgment model was constructed from these findings.

The potential of the model is as a heuristic to generate further deliberation and research in marking and as a tool for practitioners to refine their judgment policies.

Dedicated to

My husband Hugh,
children Anamaria, Gordon, Kristin,
and my mother, Merle Pratt,
who formed an incomparable
support base,

and to the memory
of my mentor
Marie I. Rasey

ACKNOWLEDGMENTS

I wish to express my gratitude to Dr. Perry Lanier, doctoral committee chairman, for his reassuring guidance, his indispensable prodding, his constructive attention to the details of the dissertation, his general wise counsel and patience; Dr. Lawrence Lezotte for his keen perceptions into the political and practical nature of the public school mission, his continuing optimism and good humor; Dr. Keith Anderson for his philosophical challenge to keep teacher accountability for pupil learning within a humanistic perspective; Dr. Ted Ward for his insistence on breadth of background and his penetrating questions to assure disciplined inquiry during the study. Special gratitude is expressed to the founders of the Institute for Research on Teaching, particularly Dr. Lee Shulman and Dr. Judith Lanier who ventured to go beyond research on teacher behavior to examine teacher thoughts and intentions. Special gratitude also to the leaders and staff of the Birmingham School District who regularly impressed upon me an understanding of the discrepancy between ideal theories and the practical realities of the classroom.

In addition, I wish to acknowledge with deepest gratitude the efforts of Sylvia Clemence and Janyce Tilmon, CPS, who devoted untold hours, energy and creative spirit to pulling this dissertation into official form. Finally, I wish to express special thanks to the five teachers who volunteered for this study, their supportive principals, and to Dr. Jerry Blanchard who encouraged the project and without whose help this research would not have been possible.

TABLE OF CONTENTS

Chapter	Page
LIST OF TABLES	viii
LIST OF FIGURES.....	x
I. INTRODUCTION.....	1
Background Research	3
Problem Statement	7
Purpose of the Study.....	8
Research Questions.....	9
Research Methodology	11
Assumptions and Limitations.....	13
Summary and Overview	14
II. REVIEW OF THE LITERATURE	16
Introduction	16
Marks and Marking	16
Pre-1920.....	18
1920-1970.....	20
Accounting for Capacity	20
Accounting for Character and Motivation.....	21
Accounting for Goals	23
1970 to the Present.....	25
Critique of the Marking Literature	28
Teacher Decision Making and Judgment	32
Teacher Effectiveness Research.....	32
Information Processing.....	34
Teacher Decision Making.....	38
Teacher Judgment.....	39
Critique of Decision-Making Research	43
Summary.....	45
III. RESEARCH SETTING, PROCEDURES AND METHODS.....	46
Introduction	46
Research Setting.....	47
Procedures	48
Data Collection.....	48
Data Organization.....	50
Data Analysis	50
Methods	51
Process Tracing.....	51
Policy Capturing	53
Utility Analysis	56
Attribution Theory	58
Stability	60
Locus.....	61

Chapter	Page
Control	62
Summary	64
IV. FINDINGS	65
Introduction	65
Composite Case.....	66
Rules	66
Procedural Rules.....	67
Contingency Rules	68
Statistical Analysis.....	68
Across the Year.....	69
Within Marking Periods	76
Verbal Analysis	80
Composite Summary.....	96
Case One - Teacher One	98
Base Data.....	98
Philosophy and Rules	98
Procedural Rules.....	98
Contingency Rules	99
Statistical Analysis.....	100
Verbal Analysis	106
Summary.....	109
Case Two—Teacher Two	110
Base Data.....	110
Philosophy and Rules	110
Procedural Rules.....	110
Contingency Rules	111
Statistical Analysis.....	113
Verbal Analysis	120
Summary.....	123
Case Three - Teacher Three	124
Base Data.....	124
Philosophy and Rules	124
Procedural Rules.....	124
Contingency Rules	125
Statistical Analysis.....	127
Verbal Analysis	132
Summary.....	135
Case Four - Teacher Four	136
Base Data.....	136
Philosophy and Rules	136
Procedural Rules.....	136
Contingency Rules	136
Statistical Analysis.....	137

Verbal Analysis	148
Summary.....	144
Chapter	Page
Case Five – Teacher Five.....	147
Base Data	147
Philosophy and Rules	147
Procedural Rules	147
Contingency Rules	148
Statistical Analysis	149
Verbal Analysis	153
Summary.....	157
V. CONCLUSIONS AND IMPLICATIONS	164
Introduction	164
Summary of Findings by Research Questions	164
Implications for Research	170
Implications for Practice	172
Summary	174
APPENDIX A	175
APPENDIX B	182
APPENDIX C	184
APPENDIX D	187
APPENDIX E	203
APPENDIX F	205
LIST OF REFERENCES	206

LIST OF TABLES

	Page
Table 4.1 Relationship between the final mark and predictor marks	70
Table 4.2 Correlations between marks and predicted marks across a year	72
Table 4.3 Controlled relationships between the final prediction and the final mark	73
Table 4.4 Composite pattern of marking averages across a year	75
Table 4.5 Distribution of marks across three marking periods--language and math.....	78
Table 4.6 Teacher attribution-utility categories	81
Table 4.7 Composite attribution-utility count and percentage	82
Table 4.8 Attribution-utility categories collapsed	87
Table 4.9 Pattern of average marks across a year--Teacher One	102
Table 4.10 Distribution of marks across three marking periods--Teacher One	104
Table 4.11 Attribution-utility percentage--Teacher One	106
Table 4.12 Cross tabulations of effort and ability--Teacher One	108
Table 4.13 Pattern of average marks across a year--Teacher Two.....	115
Table 4.14 Distribution of marks across three marking periods--Teacher Two.....	117
Table 4.15 Attribution-utility percentage--Teacher Two.....	121
Table 4.16 Cross tabulations of effort and ability--Teacher Two	123
Table 4.17 Pattern of average marks across a year--Teacher Three.....	128
Table 4.18 Distribution of marks across three marking periods--Teacher Three	131
Table 4.19 Attribution-utility percentage--Teacher Three	133

	Page
Table 4.20 Cross tabulations of effort and ability--Teacher Three	135
Table 4.21 Pattern of average marks across a year--Teacher Four	140
Table 4.22 Distribution of marks across three marking periods--Teacher Four	144
Table 4.23 Attribution-utility percentage--Teacher Four	146
Table 4.24 Cross tabulations of effort and ability--Teacher Four	151
Table 4.25 Pattern of average marks across a year--Teacher Five	155
Table 4.26 Distribution of marks across three marking periods--Teacher Five	156
Table 4.27 Attribution-Utility Percentage--Teacher Five	157
Table 4.28 Cross tabulations of effort and ability--Teacher Five.....	157

LIST OF FIGURES

	Page
Figure 4.1 Plotted relationship between the first and final marks	71
Figure 4.2 Composite marking policy (with predictions)	74
Figure 4.3 Sample record book account	77
Figure 4.4 Decision tree for marking judgment.....	85
Figure 4.5 Marking policy for Teacher One (with predictions)	101
Figure 4.6 Record book account--Teacher One	105
Figure 4.7 Marking policy for Teacher Two (with predictions)	114
Figure 4.8 Record book account--Teacher Two.....	119
Figure 4.9 Marking policy for Teacher Three (with predictions)	125
Figure 4.10 Record book account--Teacher Three	127
Figure 4.11 Marking policy for Teacher Four (with predictions)	139
Figure 4.12 Record book account--Teacher Four	143
Figure 4.13 Marking policy Teacher Five (with predictions)	150
Figure 4.14 Record book account--Teacher Five	153
Figure 4.15 Framework for marking process	161
Figure 5.1 Framework for marking process	166

CHAPTER I

INTRODUCTION

Report cards and the pupil progress marks contained therein are one of the most persistent phenomena of American educational history. Marks fulfill both a measurement and a gatekeeping function with societal consequences extending well beyond the school organization. Despite continued criticism by educators based on low reliability ratings on repeated research, despite extensive professional efforts to establish more objective substitutes, traditional teacher marks continue to dominate official records and to be the most reliable source of information on student achievement (Bejar, 1981; Evans, 1976; Lavin, 1965). Furthermore, the societal need to identify and develop outstanding talent, to assure minimal competency to all graduates, to be fiscally accountable with public funds combines with this obvious lack of an acceptable substitute to assure the use of marks for some time into the future. The process by which teachers make marking judgments, therefore, merits careful investigation.

Marks are commonly used as (1) single, summary symbols, (2) indicating achievement in some substantial segment of a student's educational enterprise, (3) given by an instructor for (4) purposes of record and report. A mark represents the teacher's judgment of pupil achievement based on such a combination of evidence as the teacher elects to use. It involves a determination of weights in value for specific products such as recitation, homework, tests, essays and for less tangible factors such as class participation, neatness in written work, mechanical correctness, industry and effort and personal agreeableness (Thorndike, 1969).

Teacher judgment is one component of general human cognitive processes. Cognitive processes (interchangeable with the terms thinking and mental processes), are those unseen phenomena of the brain which enable the

human organism to adapt behavior to the information of the environment. Within history, cognitive processes have been depicted by two major strands: memory structures to represent information; mental operations performed upon and between these memory structures (Posner, 1973). In this study, the more modern terms decision-making processes are also treated as interchangeable with cognitive processes. Judgment is defined as one type of decision or cognitive process. The study makes no attempt to distinguish between fundamental terms which are of lasting concern to the field of cognition such as static-dynamic qualities and innate learned qualities of mental operations. Instead, the study takes its lead from information processing, from the processes of selection (simplification) and inference which allow simple human memory structures to organize the complex information of a changing social environment such as the contemporary classroom.

Most histories of marks refer to the four functions of marks identified by Wrinkle (1947).

1. Administrative Functions. Marks indicate whether a student will be promoted, required to repeat or be graduated. They are used for transferring records, for judging candidates for college admission and for evaluating prospective employees.
2. Guidance Functions. Marks are used in diagnosis of special abilities or inabilities and for placement within the curriculum.
3. Information Functions. Marks are the chief means by which schools officially record pupil progress and communicate it to parents.
4. Motivation and Discipline. Marks are used to stimulate greater student effort, to determine eligibility for honors of various sorts and eligibility to play on teams.

● Feedback Function. In addition to the four functions identified by Wrinkle, a fifth function was identified. In 1967, the Yearbook of the Association

of Supervision and Curriculum Development (A.S.C.D.) related this fifth function to the theories underlying behavioral objectives and accountability. Arthur Coombs and contributing authors discussed the need for teachers and programs to have a summary evaluation component which acted as a feedback mechanism for correcting or fine tuning system errors. They argued that marks ought to fulfill such an important function or substitutes should be found (Wilhelms, 1967).

These five general functions involve some of society's biggest value questions and include the assumptions that marks can:

- Accurately measure individual achievement against an absolute standard.
- Reflect individual achievement in relation to potential.
- Reflect individual achievement in relation to group (class, school, state or national norms).
- Predict future achievement in subsequent schooling, K-12.
- Predict future achievement in college.
- Predict future job success.
- Motivate sustained or increased academic production.
- Reflect teacher effectiveness as a feedback mechanism.

Background Research

Research in several fields of education has been guided by these functions and assumptions. This study reviews the literature in two related areas; research on marks and marking and research on teacher decision making. The reviews reveal an abundance of studies on marks themselves, but an absence of systematic inquiry into the teacher judgments which underly the marking process.

Research on marks and marking has been guided by these identified functions and assumptions for more than half a century. Four separate reviews of the marking literature reveal a continuing contradiction (Evans, 1976; Kirschenbaum, Simon & Napier, 1971; Smith & Dobbins, 1959; Thorndike, 1969). On the one hand, marks are repeatedly found unsatisfactory because they fail to carry out any one of the five functions reliably on objective measures, and because they rely on subjective teacher judgment. On the other hand, marks remain the most reliable indicator of student performance (Bejar, 1981; Lavin, 1965), and recommended substitutes still rely heavily on teacher judgment. This seeming contradiction points to a discrepancy between the number of functions ascribed to marks and the number which are actually accounted for by classroom teachers.

The discrepancy may be understood by noting other themes and silences found within the marking literature. The review in Chapter II indicates that the number of functions ascribed to marks and the underlying assumptions has grown and developed with the expanding role of education in general. Research into these functions has been dominated by the measurement field which has examined each assumption singly, found it unreliable, and argued logically that the function was not being carried out well, and therefore, substitutes should be found for the whole process. Studies in this mode have relied heavily on outcome measures, usually paper and pencil tests, and have emphasized correlational strategies with limited variables under consideration. When studies have expanded to consider teacher variables, they have looked at teacher characteristics and behavior rather than teacher judgment processes. They have emphasized the static-linear nature rather than the dynamic interactive process of the classroom. The continued emphasis on studying single assumptions and limited, discrete variables, has taken the focus off the larger, complex functions and multifaceted nature of the

marking task. Yet it is from this multifaceted nature of the classroom environment that teachers select and organize the marking task.

There are notable silences within the teacher decision-making literature. Although teacher judgment is acknowledged as a basic factor in marking and in most substitutes, the judgment or organizing process has not been examined systemically. Even in the emerging literature which examines teacher planning and content choice, the marking process has not gained attention. Neither has the time and cost of the marking decision been examined in the literature, although all recommended substitutes for conventional marks make increased demands on teacher time and frequently are accompanied by pleas for revised reporting procedures, smaller classes and released time (Anderson, 1966; Thorndike, 1969). The literature is also silent regarding interaction effect between pupil level of compliance with assigned tasks and teacher judgment, although one theoretician describes the task structure of classrooms as ultimately an exchange of performance for grade (Becker, Geer, & Hughes, 1968). Finally, the emerging literature in the teacher effectiveness field, which supports the measurement of the classroom factors of immediacy and simultaneity, is silent as to how these are related to the marking process (Brophy, April 1980; Doyle, 1975).

In his review of research on marks, Robert Thorndike pointed out that marks involved conflicting values of society which made it difficult to arrive at a working consensus on marking practices. This being the case, Thorndike included an important section on marks related to the institutional culture pattern which impacts them, and he pointed out that a "modus vivendi" has typically been worked out between the tradition of marking and the rest of the institutional culture. "One who would reform the marking system of an educational institution needs first to acquire a profound understanding of the culture of that institution.

In conclusion, much of the literature on marks and marking over the past 50 years seems to have missed the mark because it has operated in an unrealistic world. It has been unrealistic in two senses. It has been insensitive to the very real limits in time, precision of judgment and skill in assessment within which the typical teacher operates. It has been largely unaware of the complex cultural pressures bearing upon instructors within the society of an educational institution and defining the bases and limits of their grading practices much more than does any psychometric theory. It is within these two limiting structures that any reform or improvement of grading practices must operate (Thorndike, 1969, p. 768).

The well-developed literature on human judgment and decision making is pertinent to the marking task. In particular, Herbert Simon's contribution in Sciences of the Artificial (1969) bears directly on the discrepancy between theoretically ascribed functions and actual teacher practices. Simon contends that the human memory is really quite limited, and in order to make an extraordinarily complex environment manageable, people construct simplified or "bounded" models of reality (Newell & Simon, 1972) and attend only to strategic or salient factors in the situation. When the environment is a changing one rather than a fixed, mechanical nature, then the rational model of decision making is idealistic. Reasonable persons cannot know all possible combinations of factors in a changing social environment such as a classroom containing 30 diverse pupils, hence, they cannot "optimize" one solution as the rational model suggests. Instead, one must plan, group, choose and simplify in order to satisfy the situation and proceed to the next problem. This becomes an exercise in information processing.

Perhaps nowhere in the teaching process is the information-processing paradigm more appropriate than in the task of marking. The process of simplifying the complex ingredients of classroom activities into symbols which have sufficient meaningfulness to be included in a summative mark is a process of design. Teachers are apparently highly selective because there are relatively few

categories in a record book, and these are representative of an entire six to eight weeks of student activity and production. Distilling one mark from approximately 10 or 20 meaningful marks in a record book which in turn were distilled from numerous activities over an extended period of weeks, is surely a remarkable act of simplification worthy of investigation. And when this distillation of classroom tasks is expected to be directly related to the functions which society ascribes to marks, it appeared important to investigate the relationship.

Problem Statement

The persistent dissatisfaction with teacher marks of student performance lies in a discrepancy between the functions ascribed to marks by society and the functions actually taken into account by teachers when judging pupil performance in the classroom context. Society has used marks as measures of academic achievement against an absolute standard (mastery), as predictors of future achievement in K-12 (diagnosis and placement), as predictors of college success (entry and credentialing), as predictors of future job success (job entry and training), as motivators for learning (reward and punishment) and as potential evaluators of teacher/program effectiveness (feedback and accountability). These functions have guided marking research. Despite repeated research findings of low reliability of marks with these functions (Evans, Kirshenbaum, Smith & Dobbins, Thorndike), marks remain the dominant system of assessing and recording pupil progress at all levels and the most influential predictor of college performance (Bejar, 1981).

The emerging research literature on teacher decision making suggests that the immediate demands of the classroom environment influence teacher decisions and planning more than theoretically based objectives or goals (Brophy, 1980;

Clark & Yinger, 1979; Joyce, 1980; Shavelson, 1980). The marking judgment, the process of selection, organization and inference regarding evidence upon which the mark will be determined, is also heavily influenced by immediate classroom demands and student characteristics. That is, teacher selection of tasks to be included in a summary mark, and the heuristics and attributions used to reach final judgment, involve a more limited and immediate set of functions than those ascribed by society in general. Since the literature on both marking and teacher decision making is silent as to marking processes, a study to determine the nature of the discrepancy and the mental process was warranted.

Purpose of the Study

The major purpose of this study was to develop an understanding of the marking judgment which engages the teacher across a school year. The principal goal was to generate a description of the thoughts, judgments and decisions of five elementary teachers during the marking task. It was hoped (1) to identify strategies and cues which determined the marking judgment and perhaps to construct a model or framework of the process from these, (2) to compare the emerging judgment factors with the functions ascribed to marks by society and (3) to generate hypotheses about the marking process which would indicate fruitful areas for future research.

An investigation of the process of marking was justified by the number of highly involved constituencies who had previously commissioned their own studies of marks, but not targeted the teacher judgment process—school districts and parents, teachers, students and the educational research community. (1) From the viewpoint of school districts and administrators, the report card remains the major communication device between schools and homes across the nation (ERS

Report, 1977). It is relied upon by parents as a personal pupil progress report (Anderson, 1966). The marking process is considered so valuable that district policies and teacher contracts specify periodic reports and often set aside paid teacher record days. (2) From the teacher viewpoint, marking student work is a task which absorbs the most significant block of professional time outside the classroom (Hilsum & Cane, 1971; Yinger, 1977) and which results in a rational system (record book) capable of explaining or justifying student marks at any given time. (3) From the student viewpoint, marks are part of a permanent record which may track students into specific skill levels or classes and which continue to be the most reliable source of achievement information for determining eventual college or job entry (Bejar, 1981). (4) Finally, from the educational research community viewpoint, the process of marking deserved to be studied in its own right as the potential site of greatest accountability where learning tasks, intentionally planned and assigned by teachers, are transformed into measurable student achievement, symbolized by a mark, on a daily, cumulative basis. In this respect, it is interesting to note that teacher education programs seldom have courses or texts pertaining to the marking process nor to its role within the larger teacher process.

Research Questions

A study of the teacher marking judgment represents a study of general human judgment. Judgment is well discussed by Johnson (1955) and Newell (1968) and summarized and reviewed by Shulman and Elstein (1975). Judgment as described by Johnson is distinguished by three thought processes: preparation, production and judgment.

The third process, judgment, may be identified as the evaluation or categorizing of an object of thought. This is logically differentiated from productive thought in that typically nothing is produced. The material is merely judged; i.e., put into one category or another. Many of the subjective analyses of thinking have included a concluding phase of hypothesis testing or verification during which the thoughts previously produced are judged. In experimental psychology, judgment is a well-developed topic, studied chiefly under the headings of psychophysics, aesthetics, attitudes and rating of personnel (Johnson, 1955, p. 51, cited in Newell, 1968).

Newell gives Johnson's definition more precision:

1. The main inputs to the process are given and available; obtaining, discovering or formulating them is not part of judgment.
2. The output is simple and well defined prior to the judgment; the judgment itself is one of a set of admissible responses; where classes or categories are given, it is usually called selection, estimation or classification.
3. The process is not simple transduction of information; judgment goes beyond the information given, adding information to the output.
4. Judgment is not simply a calculation or the application of a given rule.
5. The process of judgment concludes or occurs at the end of a more extended process.
6. The process is immediate, not extended through time with subprocesses, in which case we would refer to preparation for judgment.
7. The process is distinguished from searching, discovering or creating, as well as from musing, browsing or idly observing (Newell, 1968).

In applying these insights to an inquiry on teacher mental processes while marking pupils, Lawrence Stenhouse adds a historical dimension, retrospective generalizations, rather than the more commonly used predictive generalizations, of well-known judgment studies. Stenhouse's concern is to "map the range of

experience rather than to perceive within that range the operation of laws in the scientific sense. The utilization of history . . . works through the refinement of judgment, not the refinement of prediction" (Stenhouse, 1980).

In applying the thoughts of Stenhouse, Johnson, Newell and Shulman to the marking task, the following research questions were generated to guide the study:

1. Upon what information is the summative mark based (first, second and final mark)?
2. What cognitive process or processes make possible the formative stages (record book categories) of marking?
3. Is there a judgmental rule which explains how the input information (formative) is transformed into the output (summative), a mark?
4. If the judgmental rule yields a zone of uncertainty between any two preordained categories of judgment, what cognitive processes enable the teacher to assign a mark up or down?
5. Do identified cognitive processes form a pattern, schema or model of the marking process?
6. Do these identified teacher cognitive processes account for the five functions ascribed to marks by society in general?
7. Of the four methods of investigation used, is one superior for illuminating the marking process?

Research Methodology

The judgment literature indicates the use of numerous methodologies. Due to the nature of cognitive data and the high reliance on self-report, the field has spent considerable time critiquing its own methods in order to assure scientific validity. For teacher marking, a combination of methods allowed a system of cross-checking or corroborating evidence gathered by any one method. Hence this study used four of the most common approaches within judgment

theory: process tracing, policy capturing, attribution theory and utility theory. These approaches were grounded in the comprehensive work of Nisbett and Ross, Kahneman and Tversky, Heider and Weiner, Shulman and Elstein and Simon and Newell.

Within the decision-making literature, one term needs particular explanation for this study: policy capturing. Under ordinary circumstances, policies are understood to be the general guiding principles which determine institutional procedures and regulations. In the judgment literature, policies are private/personal principles which determine the weighting of various factors (cues) involved in a judgment. Policy capturing concentrates on identifying these cues, defining their relationship to each other, and estimating their relationship to a final judgment for the purposes of prediction and control. In this study, it results in both a composite teacher marking policy and an individual policy for each teacher depicted in Figures 4.2, 4.5, 4.7, 4.9, 4.11 and 4.13. Policy capturing is discussed in greater detail in Chapter III.

The study used classroom field data derived from a series of taped interviews (five experienced, upper elementary teachers, one suburban school district), record books, marks and predicted marks in math and language arts across one year. Quantitative and qualitative analysis focused on the formative, summative and final marks of 152 students. The distinction between these levels of marks, previously made in the research questions, is critical to any interpretation of the analysis. Formative marks are those which represent daily and/or weekly tasks within the record book columns. Summative marks are those which result from the teacher judgment at the end of an officially designated marking period and which are sent home on report cards. Final marks, sometimes called the final summative marks, are those which result from the judgment process at the end of the school year and which remain on the permanent record.

The distinction between these levels is important and can be confusing, because summative marks change roles—performing a formative function within the final mark. It is the summative marks and the final marks which serve as the major data base for quantitative, statistical procedures within this study. Formative marks and verbatim protocols form the primary data base for the qualitative methods.

The data was organized into a composite case and five cases within the schema of each teacher accompanied by his/her captured policies and strategies. These schemas supported an emerging classroom model of the teacher marking judgment.

Assumptions and Limitations

The primary assumption of the study is that teachers' thoughts guide a significant portion of the teacher marking process and that these cognitive processes, despite being unobservable, can be identified and understood. The study does not seek to prove or disprove a specific hypothesis. Since no other systematic inquiry has been undertaken on teacher marking judgments, this study is intended to be a first step in generating meaningful hypotheses for future research.

The study has important limitations. It was limited to the elementary level in order to have comparable data from the teachers studied. Although the marking research literature is heavily oriented to college and high school marks, there has been little attempt to hold grade level variables constant in state-of-the-art papers. Marking reviews jog back and forth across grade levels without much concern. From the early interviews, it became obvious that it was highly likely that elementary teachers who have students in self-contained classrooms

framed the marking judgment differently than the secondary teachers who have students for one period only. Although some implications of the study may be applicable to marking judgments at all levels, it is imperative to recognize that this study was limited to elementary teachers.

The teachers in this study are highly experienced. None has less than 14 years of teaching. This may be viewed as a limitation when generalizing to other teacher populations. However, it may also be viewed as an advantage when it is noted the average U.S. teacher is comparable in age and experience. Teachers of long experience may have refined the judgment process to the point of having very specific policies and strategies to declare.

The focal point of the study was teachers' judgment related to formative and summative marks. There are numerous areas in education which touch on marks indirectly, e.g., reward and punishment literature, student mediating literature, etc. To review each of these would have made the study unwieldy. Future studies may combine some of these areas meaningfully.

Summary and Overview

Chapter I has established that a contradiction appears in the marking literature as to the reliability and usefulness of marks and marking. The contradiction may arise from a discrepancy between the functions ascribed to marks by society and those actually accounted for by classroom teachers during the marking task. In order to create an understanding of the teacher judgment process across a year, this study of five teachers serving 152 students in a typically achieving, suburban community in Michigan was undertaken. Chapter II reviews the background literature in two related areas; research on marks and research on teacher decision making. These reviews place the problem in

historical perspective. Chapter III reviews four research methods for investigating a judgment problem which involves gathering and analyzing both qualitative and quantitative data. Process tracing, policy capturing, attribution theory and utility theory are applied to five teacher cases and one composite report. The primary mode of data collection, the structured interview, is discussed in detail. Chapter IV displays the data which teachers considered during the marking task, contains inferences made from the data and presents a model of the marking judgment process. Chapter V summarizes the findings. The model of the teacher judgment process is compared to the functions of marking as delineated in Chapter I. Implications of the study are discussed. The conclusion is summarized under the research questions which guided the study.

CHAPTER II

REVIEW OF THE LITERATURE

Introduction

Teacher judgments during the marking process have not been empirically investigated and, therefore, do not provide a body of research to review. However, studies of teacher decision-making processes and studies of marks and marking have been combined to lay the groundwork for investigating teacher thinking during the marking task. This review is divided into two major sections: Research in Marks and Marking and Research in Teacher Decision Making. Each section is followed by a brief critique, and the chapter is concluded with a summary statement.

Chapter III, Methods and Procedures, contains an additional literature review. It is limited to a review of four common methods of investigation used to examine cognitive processes, and was included because of the controversial nature of self-report data which forms the basic evidence in judgment studies. In Chapter III, the review is directly tied to the immediate application of specific methods within the marking study and, therefore, was not included as part of the general literature on marking judgments.

Marks and Marking

Research on marking parallels historical educational concerns. For the purpose of organizing this review, the literature has been divided into three phases: pre-1920 when the emphasis was on perfecting standards of measurement of pupil products usually without concern for the pupil; 1920-1970 when the

emphasis was on the learner and on increasing the functions and comprehensiveness of marks to account for pupil characteristics related to learning; 1970 to the present, when the emphasis is on teacher behavior which causes learning especially in relation to evaluation and accountability. The years used to bound a time period are characterized by particular educational developments or movements at a national level including; the Testing movement (Army Alpha & Beta tests), the Progressive Education movement (Developmental psychology), the Behavioral Objectives movement (Goals statements), the Accountability movement (Coleman Report combined with national decline in SAT achievement scores). In no sense are these dates intended to be precise cut-offs, but rather as general markers along trend lines. In fact, most movements were germinated by particular events in a preceding period. Despite national movements and expanding functions, marking practices reflect the peculiar local nature of U.S. education which blends national concerns in a local formula, frequently tied to a local tax base. And, as this study illustrates, marking continues to reflect the processes of individual, human judgment as exercised by the teacher.

The central question for research on marks has been what the mark should represent. What aspects of student performance should the mark try to characterize and in relation to what reference group should the appraisal be made? These questions set the research agenda on marks. The search for answers reveals a clear parallel with historical educational developments; a trend toward adding functions, toward more sophisticated measuring tools, toward complexity of record format. Concurrently despite increasing functions, the search reveals a strong underlying reliance on pupil achievement variables. The search also reveals that research on marks has had a limited conceptual framework which has

manipulated and correlated product and learner variables in static designs, but which has not investigated the variables of teacher and classroom.

Five reviews are especially pertinent and form the basis of this section of the study, although the section is brought up to date by reference to single studies recently completed. A seminal review by Wrinkle (1947) set forth the basic functions of marks, and all other reviews refer to Wrinkle's contribution. A review by Smith and Dobbins (1959) organized the literature between 1910 and 1957 from a historical perspective. Another by Kirschenbaum, Simon and Napier (1971) took a humanistic point of view and organized the research literature to convey the message that traditional marks were dysfunctional for a significant segment of students. This point of view was supported and updated by Evans (1976) where the literature was specifically organized around Wrinkle's four functions of marks. Lastly, Thorndike (1969) organized the literature around a traditional measurement view where marks were compared to specified frames of reference. Thorndike parted from the traditional view, however, by arguing for a new research agenda where institutional and teacher variables would be investigated. Each review was done in depth and offered an important perspective.

Pre-1920

Studies on marks have reflected the development of educational concepts. Early work reflected the long-held philosophy that knowledge existed apart from the individual. The job of teachers was to impart knowledge to students and to measure the extent to which it had been absorbed. Early research was concerned with the knowledge-product and standardizing measurement across teachers. Starch and Elliott (1912) set the pace by three simply designed studies asking approximately 100 teachers in each of three subject areas to mark a paper

on a scale of 100 points. Critics accepted the 39-point range found in the subjective area of English, but they were surprised at the range of 45 points found in geometry. The probable error in all three subjects was nearly the same; 5.4 English, 7.5 math, 7.7 history. Starch and Elliott concluded that variability was a function of the examiner and of the method of examination. With slight variations, these studies were repeated up through the 1950s with similar results. Studies by Rugg, Dearborn and Kelly between 1910 and 1915 led to the generalization that individual teachers set their own standards with a resulting variability, unreliability and inconsistency of distribution of student marks (Smith & Dobbins). Bells (1930) demonstrated that teachers regraded papers with low reliability. Tieg (1952) reported that a single teacher given the same test paper to rescore, assigned marks that differed 14 points (100-point scale) on the average from first marks.

However, studies concluding low reliability were accompanied by studies which were more optimistic. Greater reliability and less variability was achieved by merely changing from percentage points (100 categories) to letter grades (commonly five categories). Starch and Elliott supported this. Wrinkle pointed to the adoption of five symbols as the greatest single change in marking practice since 1900 and by 1939, four-fifths of elementary and secondary schools used them (Wrinkle). Less variation also resulted when serious attention was given to marking and when teachers agreed upon general governing principles (Jaggard, 1919). Common standards were the pursuit of the College Entrance Examination Board as early as 1910 (Smith & Dobbins). Studies continued to measure learning as a product to be accounted for by standards, frequencies and distribution measures. Answers to marking problems were sought by manipulating symbols and categories, and the major function of marks was to represent knowledge.

1920-1970

The years between 1920 and 1970 clearly reflect expanding functions beyond the measurement of subject matter achievement. Widespread use of the Army Alpha and Beta tests during World War I established the concept of measurable differences in capacity and went a long way toward destroying the idea that any child could learn as well as any other child if s/he tried hard enough. This concept brought an interest in individualization of instruction, increased attention to ability grouping, changes in promotion practices and wide use of standardized tests.

Accounting for Capacity. From World War I onward, the marking function was expected to take account of individual capacity. The use of standardized tests for individualization and ability grouping brought forth the problem of whether or not a full range of marks should be used in each ability group. Studies promoted various solution such as A, B and C for superior classes; B, C and D for average classes; C, D and F for slower classes. A Los Angeles committee (1926), however, recommended the whole range in each section and a Wisconsin committee (1929) recommended a symbol to denote the level at which the work was done (Smith & Dobbins). Research was inconclusive and the problem of grouping or tracking continues into present day discussions of mainstreaming and gifted education.

As compulsory attendance laws took effect and the school population increased, the interest in measuring achievement and in performance prediction became fascinating fields in their own right, separate from marking, but interestingly always reliant on marks as a point of validation (Bejar, 1981; Lavin, 1965). These fields have a voluminous research literature which is not reviewed within this study but which is referred to at a later point.

Accounting for Character and Motivation. The impact of testing on marking was strengthened by the impact of the Progressive Education movement undergirded by developmental psychology. The movement was gathering momentum prior to World War I. The Progressive Education Association officially stated: "The aim of Progressive Education is the freest and fullest development of the individual, based upon the scientific study of his mental, physical, spiritual and social characteristics and needs" (Cremin, 1951, p. 240). Cremin characterized the movement as a many-sided effort to use the schools to improve the lives of individuals by: first, broadening the program and function of the school to include direct concern for health, vocation and quality of family and community life; second, by applying scientific, pedagogical research in the classroom; third, by tailoring instruction to the different kinds and classes of children who were being brought within the purview of the school. This broadening viewpoint grew in momentum and influenced most educational procedures including pupil records and marks (Cremin).

Whereas the testing movement greatly influenced grouping and promotion practices related to marks, the Progressive Education movement emphasized pupil characteristics which were related to level of development and motivation. Ethical character, citizenship, responsibility, industriousness, worthy use of leisure time—these factors were considered to be important and related to learning. Report card formats responded by carrying checklists which supplemented symbols. Studies showed that teachers considered effort, attitude and other factors besides achievement and that the term "achievement" had many component parts such as accuracy, mastery, regularity and application. The use of rating scales and checklists in both academic (cognitive) and nonacademic (affective) areas offered a basis for marking which was behaviorally defined. Freyd (1923) found rating scales yielded more reliability between raters and

between the same rater over time than did marks. By 1939, 87 percent of elementary report forms listed character traits (Wrinkle, 1947).

Evaluation in this manner, however, was often difficult to explain to parents because of overlapping terms such as "reliability," "dependability" and "responsibility" (Wrinkle). Gradually, parent conferences were instituted to supplement both symbols and checklists. Conferences were found especially useful when one teacher had a class of youngsters for the whole day, but the time commitment to this process was heavy. Evidence indicates that supplements to symbols developed steadily at the elementary level with a clearly emerging relationship between wealth of the school district and extent of the supplements (Thorndike, 1969). The marking function was expected to account for pupil traits which enhanced or detracted from school learning.

The 1930s marking studies reflected a growing criticism of marks on the part of the Progressive Education movement. The major thrust of criticism was directed toward the competitive nature of marking which was believed to seriously harm student motivation especially students from poor home backgrounds (Wrinkle). Research studies were directed toward the relationship between motivation and grades. Marks as major incentives were heavily criticized in the belief that they often led to undesirable behaviors such as cheating, superficial learning and complacent attitudes. Some educators thought they encouraged complacency and/or fostered fear (Smith & Dobbins). Tiegs (1931), however, found that with intermediate pupils, 90 percent tried harder because of good marks and 97 percent because of poor marks. Fay (1937) reported that A students do better when told their marks, B students somewhat better and C students only slightly better (Smith & Dobbins). Although feelings ran high on this topic, research was inconclusive. More recently, research indicates that usually the better students experience greater motivation from grades (McDavid, 1959;

Phillips, 1962) and down-graded students often continue to fail (Chansky, 1962). A number of researchers have observed that teacher marks are often skewed to the high end of the curve, and that this skewing is to be expected in selected groups (Smith & Dobbins), but skewing was not examined as a motivational technique. Research on marks and motivation, spanning the years between 1930 and 1970, can be summarized by saying that different students respond to grades differently. Wrinkle, however, concluded that of all the functions which marks were expected to carry out, the motivational function was actually the "only one they served with any considerable degree of effectiveness." Heavy criticism of marks continued through the reviews of Kirschenbaum and Evans with two major effects; first, it firmly established the motivational function of marks; second, it promoted the search for substitutes.

The 1940s movement to Behavioral Objectives evolved from the search of Progressive Education for a substitute for marks. It was also encouraged by the realization that no one could be sure what a single mark meant unless it represented the measurement of a single identified value. If the change from percentages to five symbols was a hallmark in the history of marking, the movement to objectives was surely as important. However, despite its original promise as a substitute for marks, its eventual impact was indirect. Its profound effect was to reground the marking function in the curriculum rather than in the competing functions of selection, prediction, reporting or motivation. Aside from a few marking experiments, Behavioral Objectives changed the process of marking rather than marks themselves.

Accounting for Goals. Research on marks and marking since 1940 reflects the growing momentum of the Behavioral Objectives movement and its absorption by the Accountability movement. Marking studies indicated a growing practice of attempting to evaluate much more than subject-matter achievement (Traxler,

1957), however, the general terms of "citizenship" and "responsibility" were not satisfactory due to overlapping meanings. The improvement of written comments became a distinct goal and borrowed much from Traxler in "The Nature and Use of Anecdotal Records" (Traxler, 1939; Wrinkle, 1947).

The Ten-Year Study at Colorado State College of Education shows the emerging relationship between marking and Behavioral Objectives. In 1929 the Colorado Campus Research Laboratory announced it was going out of the A B C D E marking business:

Over the next ten years we made almost every mistake a school could make in our efforts to improve our marking and reporting practices. In rapid succession we developed and discarded innumerable detailed evaluation report forms, checklists and scale-type reports. We juggled symbols--S U, H S U, H M L, and others. We accumulated thick files of anecdotal records. We tried informal-letter reports. For a time we abandoned all forms of written reporting and substituted parent-teacher conferences. We constructed elaborate cumulative record forms. We emphasized student self-evaluation. We developed still other detailed report forms. And in every direction we went we came out at the same spot. If it was good, it took too much time; it wasn't practical; it wouldn't work in the public schools. And our job as a research-laboratory school was to work out not only something we could use; it had to be equally useful for Yuman or Yampa or Teaneck or Tacoma (Wrinkle, p. 4).

In the 1940s, Wrinkle looked back and wrote:

From the beginning of our work we recognized the importance of outcomes other than those specifically associated with subject matter. We considered self-directiveness and social adjustment and others of such importance that we built checklists, scales, self-evaluation forms, and other devices to evaluate them. But it did not dawn on us until about 1938 that if the end product of a part of a youngster's experience could and should be measured in terms of what he does, the end product of all educative experience is the modification of behavior of the learner. That is why we have schools. If a learning experience does not result in a modification of the way the learner behaves, he has not really learned anything of real value (Wrinkle, 1947, p. 4).

Behavioral Objectives eventually became a powerful center of controversy. Bloom (1956) created a taxonomy for classifying objectives which fueled educational research for years. He went on to establish percentages of mastery within objectives and eventually to assert that aptitudes were alterable, that all students could learn most objectives given sufficient time and appropriate instruction (Bloom, 1976). The tendency of some parts of the measurement field to define smaller and smaller units of behavior, however, was viewed with growing alarm even by such advocates and seminal thinkers as Ralph Tyler (Shane, 1973) and James Popham (1972). Tyler repeatedly cautioned that human nature was better portrayed by higher or more general objectives than by the specific ones currently used (Tyler, 1950). Popham noted that some important goals were unassessible and in some classes the proportion of nonmeasurable goals might be smaller than others.

By the 1960s, it was clear the Behavioral Objectives would not be an acceptable substitute for marks. The list of Behavioral Objectives grew so long that it had to be classified into the domains of cognitive, affective and psychomotor. Lists of objectives on report cards represented the height of the expansion of the format and function of report cards. Following this attempt at substitution for marks, no new solutions have been offered within the marking literature, although minimal competency tests have been offered as a product bottom line.

1970 to the Present

Prior to 1970, commissioned by the federal legislature to perform the first massive assessment of public education, Coleman published his correlational research Equality of Opportunity (Coleman, 1966). His conclusion was that despite massive compensatory spending programs sponsored by the Civil Rights

Movement, the schools were not fulfilling the job of equalizing the outcomes of all students. In 1969 heeding Coleman's message, Leon Lessinger articulated the government's accountability theory in education whereby schools were to be responsible for actual learning outcomes not just for providing opportunities (Lessinger, 1969, 1975).

Accountability for outcomes following 1970, was fueled by a swelling tide of public conservatism, fiscal and otherwise, and the startling fact, emanating from the Scholastic Aptitude testing service, that the achievement scores of college bound students had declined steadily since 1963 (S.A.T. Panel Report, 1977).

The period of expansion of marking functions paralleling the period of curriculum expansion, experienced the strong impact of evaluative measurements and publication of the results. The lack of success of expensive compensatory education programs, the declining achievement scores and the general swelling of conservative sentiment caused educators to debate which of the many functions which had accrued to education over the years were truly their responsibility. Educational research since 1970 reflects an evaluative and reflective mode and a reemergence of educational historians searching the past for patterns of development and success across the years (Broudy, 1972; Cremin, 1961, 1965; Tyack, 1980).

In particular, Cremin documents the formal demise of Progressive Education in the 1950s (Cremin, 1961), although others believe it reemerged briefly with the Civil Rights Movement (N.E.A., 1974).

The marking literature after 1970 is characterized by a drop off in research studies, by literature reviews which are forced to refer to studies at least twenty years old, and by very specific investigations into declining achievement scores related to a phenomena called "grade inflation." The first two

characteristics are self-evident in the literature reviews, but grade inflation is of particular interest. Evans (1976) reported that during the 1950s and 1960s, aptitude test scores increased, but grade distributions remained unchanged. Aiken (1963) found that average grades remained unchanged despite rising S.A.T. scores. In 1972 Baird and Feister analyzed data from large samples and concluded:

This study confirms the earlier research . . . which indicated that faculty members, at least collectively, prefer or are committed to a certain distribution of grades. Thus, faculties show an "adaption level" by awarding, on the average, about the same average and distribution of grades, whether their current students were brighter or duller than last year's (Baird & Feister, 1972, p. 440).

However, toward the end of the 1960s, this situation reversed. Aptitude scores started to drop and average grades awarded to students started increasing. This phenomenon has been carefully reported in the S.A.T. Panel Report (1977). Termed "grade inflation," this trend has occurred at both the college and high school level (Ferguson & Maxey, 1975), but the rate of grade inflation has diminished since 1974 (Bejar, 1981). According to Evans, the resulting situation indicates that grades reflect different achievement levels at various time periods because different levels of competition prevail. It also indicates that within a limited range, teacher marks reflect changing educational values with allowance for lag time.

By 1974, a survey by N.E.A. indicated that the five-symbol system of reporting pupil progress was dominant in the majority of the nation's schools. At the fourth-grade level, approximately seventy percent of the schools also used a supplement such as parent conferences. Following the fourth grade, supplementary information declines in use (N.E.A., 1974). However, the use of grades for promotion and eventual college entry remains the dominant pattern

through the 70s to the present, and pupil achievement is the primary for A B C symbols.

In summary, the literature on marks and marking has expanded from theoretically representing one function—a quantity of knowledge—to representing five very complex functions. Since the 1960s the expansion appears to have tapered off. The criticism by Progressive Education that marks did not encompass enough of the functions of learning has been replaced by the criticism of the conservatives that marks are attempting to reflect too much, hence, "grade inflation." Although there is no evidence of a contraction of functions, many evaluative studies have been suggested. The present study of teacher cognitive processes is representative of this trend

Critique of the Marking Literature

The thoughts of Joseph Schwab on curriculum are especially pertinent to marking. In 1969 he wrote of the field of curriculum that it was "moribund," unable by its present methods and principles to continue its work and desperately in need of new and more effective principles. Signs of such a crisis include a flight from the subject of the field including a sign of "marked perseveration, a repetition of old and familiar knowledge in new languages which add little or nothing to the old meanings as embodied in the older and familiar language or repetition of old and familiar formulations by way of criticisms or minor additions and modifications" (Schwab, 1969, p. 4). Schwab could have been describing research on marks since the 1960s. He went on, however, to suggest that the problem resulted due to an unexamined reliance on theory in an area where theory is partly inappropriate and partly inadequate to the task.

Schwab argued that curriculum was a practical rather than a theoretical art. Practical arts begin with the requirement that existing institutions and existing practices be preserved and altered piecemeal, not dismantled and replaced, because the functioning of the whole must remain current. That is, the practical is concerned with the effects of the proposed pattern of change through time in order that they retain coherence and relevance to one another (Schwab).

Research on marking has been limited by a narrow conceptual framework which has (1) concentrated on product and learner variables, (2) concentrated on static designs and (3) concentrated at an idealistic level of what ought to be rather than the realistic, practical classroom as it is. In this respect, research on marking reflects similar problems to curriculum, the problems of psychological research applied to education.

The conceptual framework of marking has been limited by a concentration on product and learner variables and by a neglect of the judgment processes of teachers. Yet, beginning with the work of Starch and Elliott in 1912, teacher judgment has been acknowledged as the crucial element in variation. But Starch and Elliott did not study teacher judgment in relation to human judgment in general. They preferred to work with product standards, marking frequencies and distributions. By 1947, Wrinkle stated that the "assumption that anyone except the person who gives a mark can look at it and tell with any degree of accuracy what it means is the No. 1 fallacy involved in the use of the conventional marking system" (Wrinkle, p. 35). But Wrinkle did not study the cognitive processes which allow teachers to combine elements into a judgment. Instead he attempted to solve the problem by defining behavior into small easily observable behavioral units. The marking literature contains several references to teachers' tendency to skew marks upward, but there are no systematic studies of teachers awareness of or rationale for this phenomenon. Hence the conceptual framework has remained limited and repetitious.

Lavin, in a review of 300 empirical predictive studies of achievement, noted that research had not studied the teacher grading process. "A student's grade is more than something that characterizes him as does his score on a personality inventory or an intelligence test; . . . rather a grade should be viewed as a function of the interaction between student and teacher It is clear that if we want to predict a grade, we must know something not only about the student, but about the teacher as well." Lavin argued in 1965 for a study of the subjective factors in teachers' grading practices, even though they were harder to define and to measure reliably (Lavin, 1965). To date, that study has been neglected.

Cronbach argued in a similar fashion in 1975. Noting that the original role of the scientist to observe, had been neglected in favor of controlled experimentation, he pointed out the severe limitations of experimental or individual psychology to detect interactions of the sort Lavin called for. He attributed this limitation to the need in physical science to have a fixed reality, a controlled variable or a limited range of situations. "Rarely is a social or behavioral phenomenon isolated enough to have this steady process property" (Cronbach, 1975, p. 682). Cronbach called for a return to a disciplined observation of uncontrolled conditions, of personal characteristics and of events that occurred during treatment and measurement. One reasonable aspiration of psychology is to assess local events accurately in order to improve short-run control, short-run empiricism. Cronbach's call for careful description has been answered by numerous ethnography studies of classroom behavioral processes, but descriptions of cognitive processes have been slower to appear. Hence the mental processes which help teachers to simplify the complex environments of the classroom have essentially been neglected.

Thorndike's review of the marking literature supports both Lavin and Cronbach by concluding that any future improvement in marking practices had to be sensitive to the very real limits in time and precision of judgment within which the typical teacher operates. He stated further that research had been largely unaware of the complex cultural pressures which bear upon an instructor much more than any psychometric theory. In this respect, he pointed to the "modus vivendi" which has typically been worked out between the traditions of marking and the rest of the institutional culture (Thorndike, p. 765), a "modus vivendi" which is typical of the practical arts but often neglected by theoreticians.

The conceptual framework has been further limited by static research designs which have allowed the assessment of pupil progress to be dominated by one final (summative) mark, paper or test. Teachers, however, give several marks during a year and usually these are based on numerous task assignments or tests. Few studies have examined the impact of performance patterns. In these few which have, a relatively consistent pattern of evaluation has been found: continuously high performance > ascending performance > descending (or random) performance > continuously low performance (Ryan & Levine, 1981). Other linear experiments have investigated whether expectancies reflect a primary or a recency effect with final performance being dominant. A current study by Ryan and Levine found a modified recency model which assumes that the weight assigned to final performance varies inversely with the difference between final and next to final performance to be most reflective of observers evaluative processes. Predictions were different from evaluations with subjects less willing to make a long-term bet on a "late bloomer" than on one who showed generally high or consistently improving performance. These sequential studies of patterns within marking show promise and need replication in natural settings.

The documented limitations of the conceptual framework in marking research are strikingly similar to the situation in curriculum described in 1969 by Joseph Schwab. Marking functions may lend themselves much more to practical arts than to theoretical and ideal circumstances. The ability of teachers to consider a practical combination of functions in a single symbol is best discussed through the literature on teacher decision making which follows.

Teacher Decision Making and Judgment

Teacher decision making is a relatively new field. It emerged from the field on Teacher Effectiveness in response to a growing criticism that despite extensive research, few consistent relationships between teacher variables and effectiveness criteria had been established (Doyle, 1975; Dunkin & Biddle, 1974; Getzels & Jackson, 1963; Medley & Mitzel, 1959). In addition, the field was characterized by a growing discrepancy between theoretically based prescriptions and actual classroom practice (Doyle, 1975). Teacher decision-making studies were an attempt on the part of researchers to get beyond the variables of observable behavior to the teacher mental processes which select tasks and guide classroom behavior. Such inquiry would shed light on the growing discrepancy between theory and practice.

Teacher Effectiveness Research

The literature on teacher decision making is limited, but its roots are well established in the teacher effectiveness literature. The bulk of effectiveness research concentrated on four classifications of variables; presage variables (teacher characteristics such as age, sex, social class and training), context variables (grade level, subject matter, size of class and community), process

variables (teaching method and style, talking and questioning strategies) and product variables (achievement and tests). In contrast to the marking literature which focused largely on product and learner variables (aptitude and intelligence), teacher effectiveness research concentrated on process variables in an attempt to isolate generic skills which would be taught in teacher training institutions. Empirical research, manipulating or counting these process variables in controlled settings, was extensive and has been well reviewed (Dunkin & Biddle, 1974; Gage, 1978; Medley, 1977; Rosenshine, 1976).

The teacher-effectiveness research contributed greatly to the identification, specificity and frequency of teacher actions in the classroom (questioning, reinforcing, wait-time, acknowledging students, task orientation, praise, absence of criticism; Gage, 1978; Medley, 1979; Rosenshine, 1979). Gage, sifting through several hundred variables in teacher behavior and classroom activity, summarized the implications of the research to say that teachers should organize and manage classes so as to optimize the concept of "academic learning time"—time in which pupils are actively engaged in their academic tasks (Gage, 1978, p. 39, 40). To this end, Gage listed seven "teacher-should" statements which he felt were easily inferred from research data. Hence the search for overt acts of teachers which were related to student achievement bore fruit, and recently Rosenshine (1979) described process-product research as "alive, well, and continuing."

Nevertheless, teacher-effectiveness research came under continued criticism much of it from within its own research ranks. The process-product approach was limited conceptually (Brophy, 1980; Doyle, 1975; Shavelson, 1980) primarily because it had not taken account of teacher thoughts and intentions in organizing the classroom. Brophy, prominent in the field, stated that effectiveness research had been virtually silent on the topic of teachers' thoughts

while engaged in the act of teaching. He attributed this silence to the pervasive influence of behaviorism on American social science research which tended to look upon thoughts as "mere epiphenomena accompanying behavior" and which tended to focus on observable teacher skills rather than decision making. However, Brophy stated that this position had softened even among serious behaviorists like Bandura (1977) and Meichenbaum (1977) who had recently stressed the role of thinking (self-talk, verbal behavior) in directing behavior (Brophy, April 1980). In his latest paper, Recent Research on Teaching, Brophy noted the limits of teacher research during the 70s and predicts a shift from studying teacher effects as measured on end-of-the-year achievement tests to studying immediate or at least short-term outcomes of instruction, with increasing attention to the performance demands that different teaching behaviors and decisions impose on students (Brophy, November 1980, p. 20).

Others support Brophy's point of view, finding many classroom and teacher problems which are not addressed by the process-product approach. Shavelson notes that recommendations about increasing the frequency of a teaching act say nothing about when to act (N.I.E., 1975; Shavelson, 1980). Hence a teacher may possess a full range of teaching skills but not have strategies to determine when they are appropriate. In a comprehensive paper on the shifting paradigms of research, Doyle had predicted a shift away from the conceptual limitations of the process-product approach as early as 1975, although he foresaw more emphasis on the environmental press of classroom demands (ecological) which establish limits to the range of response options available to the teacher (Doyle, 1977).

Information Processing

The basis for a study of teacher thoughts has been well laid at the Institute for Research on Teaching. The model for the institute was one of

Clinical Information Processing, presented at the National Conference on Studies in Teaching by Lee Shulman (N.I.E., 1975). The focal point of the model and of the institute remains cognitive functioning, or the mental life of the teacher. How do teachers process the information of the classroom to create an environment which promotes achievement? The information of the classroom includes student characteristics, student records, subject matter, curriculum resources, organizational structure of the school, the accountability system, etc. Each of these serves as a set of "cues" from which the teacher makes diagnostic judgments and plans prescriptive actions. The panel which supported the general clinical model was anxious that it not be a sterile focus, and they made a commitment to whatever forms of disciplined inquiry seemed appropriate to the research problem and educational setting under investigation. Hence the model has inspired research into a variety of teacher decisions, and it has encouraged the borrowing of conceptual research frameworks from related fields such as cognitive psychology and anthropology.

The theoretical grounding for the information-processing approach came primarily from cognitive psychology with meaningful elaboration by learning psychology. In particular, information processing acknowledged the fact that the ability of people to process all of the information in their complex environment was limited by the capacity of short-term memory (Newell & Simon, 1972; Simon, 1969). Human decision making and judgment was able to overcome some of the limitations by using the identified processes of simplifying and inferring. Simplifying strategies such as "chunking" information into more abstract units, increases the amount of information processed (Miller, 1956). Simon specified seven chunks plus or minus two while Craik lowered this estimate to three chunks,

plus or minus one (1971).¹ People also selectively perceive and interpret portions of the available information and construct a simplified model of reality (Newell & Simon, 1972). By "bounding" the system within which they are operating, decision-makers do not have to deal with the real world in its totality but rather with only those "strategic factors" in a situation, once again reducing the demands of the environment (Simon, 1957). In this context, Simon argued against the rational decision models which held that all possible alternatives and their outcomes should be examined before a decision is made. Instead, Simon contended that administrative man, as distinguished from the ideal economic man, looks for a course of action that is satisfactory rather than optimal. Along with March, Simon elaborated the notions of "satisficing" and "bounded rationality" as depicting the reasonable thinking of man in a complex, confusing environment (March & Simon, 1957; Yinger, 1977).

A second mental strategy underlying information processing is inferring or going beyond available data. As discussed in the cognitive literature, the most common tools of inferencing include the creation of knowledge structures (schemas, beliefs), the availability heuristic and the representative heuristic (Nisbett & Ross). Kahneman and Tversky (1973) noted that decisions often required people to judge the relative frequency of particular objects or the likelihood of particular events. In so judging, they may be influenced by the availability of the objects or events, that is, by their accessibility in memory processes. To the extent that this availability is associated with objective frequency, it is a useful judgment tool. However, cognitive psychologists have

¹Chunking is pertinent to this study when considering that arguments within marking research frequently revolved around having three, five or seven symbols in the system (Smith & Dobbins, 1959).

found it subject to typical biases. Kahneman and Tversky described a second tool which aids people in decisions and which they termed the representativeness heuristic. People assess the degree to which the salient features of the object are representative of, or similar to, the features presumed to be characteristic of the category. Although this heuristic is useful, indeed an absolutely essential tool, it too is subject to typical problems—the most common being the neglect of important information contained in relevant base rates (Nisbett & Ross). Nisbett and Ross contend that people's understanding of the flow of social events probably depends less on specific use of tools than on such procedures as a rich store of general knowledge of objects, people, events and their relationships. This knowledge is represented as beliefs or theories or schemas. Such knowledge structures provide an interpretive framework which helps supplement information and resolve ambiguity. The concept is interesting to psychologists, and Nisbett and Ross refer to a growing list of terms including "frames," "scripts," "prototypes" and the more general term "schemas" (for a detailed discussion of these structures, one is referred to Nisbett & Ross, Human Inference).

Cognitive psychology has studied these mental processes in applied settings in order to understand people's actions in certain complex environments. Given the growing awareness of the complexity of the classroom environment, given the increasing functions of education and, given the limited predictable relationships found between variables in previous educational research, it seems natural that researchers in teacher effectiveness have begun to join fields to explore the mental processes which guide applied behavior. To date, these studies within education are limited, informative and growing in number. Most have concentrated on teacher choice of alternative curriculum, teacher allocation of time, teacher interactive decision, teacher conception of subject and teacher planning. It is notable that none have been undertaken on teacher judgment during the marking task.

Teacher Decision Making

Since 1970 there have been approximately forty studies of teacher decision making. Reviews of the teacher thinking literature are organized around the categories of planning, judgment, interactive decision making and others (Brophy, April 1980; Clark & Yinger, 1978; Joyce, 1980; Shavelson, 1980). Since each review examines most of the same studies, this review will summarize the trends and conclusions, referring only to specific studies under the category of judgment.

Most of the planning studies have been curious about the extent to which the model of rational curriculum planning is used in teacher plans. This applied model was first proposed by Tyler (1950) and later elaborated by Taba (1962) and Popham and Baker (1970). It recommended four essential steps: (1) Specify objectives, (2) Select learning activities, (3) Organize learning activities and (4) Specify evaluation procedures (Clark & Yinger, 1978).

The findings of Zahorik and Taylor, using questionnaires and rankings, were that teachers did not begin their planning with objectives, but rather with content (subject matter) or resources. Contrary to the theorists' ideas, procedures for evaluation were an issue of minor importance (Taylor, 1970; Zahorik, 1970). Following 1975 research on planning focused on simulated planning situations rather than rankings (Peterson, Marx & Clark, 1978; Morine, 1976). Simulated cases in laboratory settings then gave way to classroom field studies (Morine, 1975; Smith and Sendelbach, 1979; Yinger, 1976). The latter relied heavily on self-report methods, especially verbal analysis of "think aloud" sessions or written plans and stimulated recall procedures. The general conclusions of the planning research were that teachers' plans are unique, that teachers focus first on content (subject-matter), that they spend much of their

planning time on activities or instructional tasks and that they seldom refer to behavioral objectives.

The planning literature joined the interactive decision-making literature on the importance of maintaining an activity flow during class periods. To this end, plans are routinized to minimize conscious decision making during interactive teaching (Clark & Yinger, 1979; Joyce, 1979; Mackay & Marland, 1978; Morine & Dershimer, 1978). Conscious decision making, therefore, usually arises when the teaching routine is not going as planned.

Joyce's review of planning and interactive decision making supported the notion that the most powerful decisions were made early in the year. These decisions involved the selection of instructional materials and the development of a flow of activities which enabled children to approach tasks which were embedded in the material. This flow limited the potential stimuli to which students and teacher responded, establishing routines and parameters of on-and-off task behavior—this became the basis of information processing and fine tuning. Once this flow was set up, teachers rarely made decisions which changed directions. Instead, their concerns divided between pupil achievement (appropriate response to content of tasks) and involvement (maintenance of the on-the-task behavior). In other words, said Joyce, teachers do not think as instructional designers do, continuously selecting new methods and ways of reaching children, but rather, they work within a general design set up early in the year (Joyce, 1979).

Teacher Judgment

Along with planning and interactive decision making, judgment has been investigated. Reviewing judgment studies is complicated by the lack of distinction between the terms judgment, planning, prediction and decision

making. For example, planning and judgment are frequently lumped together in literature reviews, although planning is generally defined as a beginning act or preactive. Judgment, on the other hand, is located at the end of a sequence of operations, as a post-active operation, as the final "assignment of an object to a small number of specified categories" (Johnson, 1972). Prediction is more closely associated with planning while evaluation is associated with judgment. Within the teacher studies to date, these distinctions have not been clear, although Einhorn and Kleinmuntz recognized the problem and suggested "that the distinction between judgment and choice be maintained and sharpened" (Einhorn, 1979). Therefore, when is judgment most important in teaching? Clark answered that "teacher judgment plays an important part in predicting student cognitive and affective achievement, predicting teacher's use of instructional moves, teacher planning and teacher selection of instructional activities" (Clark, 1978). Clark's definition is obviously broad and points to the lack of clear classification of judgment studies, yet it depicted the newness of this focus of research for which Shavelson is now urging a taxonomy (Shavelson, 1980).

Teacher judgment studies, as classified in the literature reviews of Clark, Shavelson and Brophy, are limited to approximately ten studies. Anderson (1977) studied the judgment policies of 164 high school teachers in regard to 36 different hypothetical descriptions. Teachers were asked to rate the descriptions on a nine-point scale from very poor to outstanding. In addition, teachers were asked to rate each characteristic separately and to rank order them. Conclusions were for greater consistency in areas ranked least important, such as establishment of objectives and homework requirements and inconsistency in areas ranked most important, such as knowledge of subject and fairness in grading. The general conclusion was that teachers may base their decisions on a policy different from the one they report using.

In a frequently quoted study, Shavelson et al. (1977) investigated the sensitivity of teachers to the reliability of information received and willingness to revise judgments when presented with new information. This experiment involved 164 graduate students in education with a hypothetical case study. The major finding was that teachers are sensitive to reliability of information and will revise estimates. This finding contradicts the major findings of Kahneman and Tversky that people in general were neither sensitive to reliability of information nor willing to revise. In a related study involving anchoring, Joyce et al. asked 10 teachers to perform pupil sorts at repeated times during the school year and to predict end-of-the-year reading achievement. In sorting, the major cues were student personality and involvement. Other cues included ability and achievement. The most obvious finding of the study was that teacher prediction of reading did not differ substantially between September and November, even though teachers had much more information. This stability of judgment contrasts with the findings of Shavelson et al.

A study by Clark et al. (1978) attempted to identify the cues within language arts activities which caused teachers to judge them useful. Clark asked 14 teachers to rate 26 language arts activities. He then asked them to reexamine each activity rated highly and list the attractive features. This was repeated for activities rated low. Student motivation and involvement were mentioned most frequently followed by features thought to influence cognitive and affective outcomes. Level of estimated difficulty of the subject came third.

Marx (1978) studied the judgment cues teachers used in predicting cognitive and affective outcomes. Twelve teachers taught a series of three social studies lessons to groups of eight junior high students in a laboratory setting. After each session, the teachers made predictions of the rank order of students and described student behavior or other cues which caused the prediction. Marx found that participation was the most frequently used cue.

In a subsection on judgment, in his review of teacher cognitive activities, Brophy (April 1980) referred to many of the same studies. He referred back to expectancy studies which indicated that teachers can and sometimes will make predictions about student achievement on the basis of such factors as race, handwriting neatness or physical attractiveness (Brophy & Good, 1974). He was also intrigued by the tendency of first impressions to "anchor" later perceptions as noted in Joyce's study. This anchoring appears in documented studies where students who do well early and then tail off tend to receive higher predictions of future achievement than students who ultimately earn the same average score (Brophy, April 1980). This last finding of "anchoring" is questioned by the conclusions of Ryan and Levine, mentioned earlier which indicated that people use different sequencing cues for prediction than for evaluation, anchoring being more closely related to prediction and recency to evaluation or final judgment.

On the whole, Brophy reports that analyses indicated that teachers based their estimates of student achievement and classroom conduct on the cues most relevant . . . reading achievement on prior reading, math predictions on prior math, and behavior on previous classroom conduct. A study by Willis (1972) supported this. Willis asked first grade teachers to predict end-of-the-year achievement during the first week of class. Teachers early rankings correlated highly with end-of-semester rankings. Interviews in regard to criteria of judgment revealed self-confidence, participation in class, general maturity and ability to work well independently. After a few weeks of experience with the students, even more weight was given to observed classroom performance in reading. This agreed with Shavelson's results and similar findings were reported by Long and Henderson (1972).

Brophy concluded that compared to the work on teacher planning, "teacher judgment has yet to 'jell' as a subfield with structure and direction."

Critique of Decision-Making Research

All reviews point out the newness of the field and the lack of studies to critique. Shavelson's review (1980) pointed out that the planning studies were not sufficient. Most of the research was descriptive. Little was known about how and why activities were constructed or under what conditions they were used. This might help explain the contradiction between simulated studies which have indicated that teachers do use objectives (when probed) and naturalistic studies which find they seldom use objectives. Brophy's critical comments also called for going beyond descriptive studies to identify aspects of quality in teacher planning, ways to conceptualize, measure and document differences in quality.

The findings of the planning studies reveal remarkable agreement. The apparent lack of usefulness of objectives needs more investigation before prescriptions are made in teacher education. Most samples are small, most studies unreplicated.

Aside from the obvious newness of the field, the outstanding characteristic of the decision-making research is the general absence of outcome relationships. What are the teachers' thoughts and intentions about final achievement, given a year of instructional time from a student's life? How is the effectiveness of planned tasks evaluated? Is "maintaining activity flow" related to Gage's admonition to optimize "academic learning time"?

Doyle's work is extremely pertinent to the lack of established relationship between planned tasks and outcome. He suggested that the formal task structure of the classroom is defined as an exchange of performance for grades. Becker, Geer and Hughes (1968) also contended that this "exchange of performance for grades is, formally and institutionally, what the class is about." Schellenberg (1965) described the task structure as an exchange of performance for status. Student differential attentiveness is related to what the student perceives is

important to the teacher and the exchange process (Doyle, 1975). This exchange is obviously related to the activity flow of the classroom. Nevertheless, the marking process is never mentioned in the teacher decision-making literature, although it provides a promising site from which to examine teacher thoughts about outcome.

A second characteristics of the literature is its concentration on identifying cues which account for static decisions. The research techniques concentrate on one-time products (achievement tests) or one-time decisions rather than on a series of decisions over time, a progression. Resnick describes this as a weakness of cognitive psychology in general. "Cognitive psychology has been largely concerned with describing 'moments' of performance or 'states' of understanding. It can be characterized as a 'transparent snapshot' psychology, in which mental processes are depicted at a given point in time . . . (In contrast,) learning theorists record 'movies' rather than snapshots but continue to seek primarily overt behavior." Resnick seeks a blending of the strengths of each specialty for the purpose of building a psychology of instruction (Resnick, 1981). Resnick's concerns bear directly upon the decision-making literature which has reflected much of cognitive psychology's development

The decision-making literature also lacks clarity of definition. In particular, the term judgment is used loosely to cover all cases where "cues" are used in arriving at a decision. Yet some studies indicate that different strategies may be used for evaluating than for predicting, for marking than for planning. Shavelson's suggestion for a taxonomy is especially worthy in this context. Work by Einhorn, Kleinmuntz and Kleinmuntz on linear regression and process tracing in judgment studies concluded a section by stating "in any event, more work is necessary to clearly distinguish judgment from choice and the processes that may be invoked by each" (Einhorn, 1979).

Given a similar discrepancy between increasing theoretical prescriptions for teacher effectiveness and for teacher marking and given the limited application of these prescriptions in the classroom, an investigation into teacher mental processes during marking yielded useful insights.

Summary

This literature review has attempted to place the study of teacher judgment during the marking process into the framework of pertinent related research. The obvious conclusion from the review is that historically the functions of marks have increased over the years. However, despite research and new prescriptions, teacher responsiveness to this increase has been limited, causing a continuous dissatisfaction. A related conclusion, equally obvious, is that teacher judgment processes have been neglected despite their acknowledged role in determining marks. This study was a first step in connecting these areas and exploring this neglected site of research.

CHAPTER III

RESEARCH SETTING, PROCEDURES AND METHODS

Introduction

Conventional teacher marks, A B C D E, symbolizing pupil progress are likely to dominate official records for some time to come. This study was designed to capture the judgment processes of teachers during marking across one school year. It addressed the following research questions:

1. Upon what information is the summative mark (first, second and final based)?
2. What cognitive processes make possible the formative stages (record book categories) of marking?
3. Is there a judgmental rule which explains how the input information (formative) is transformed into the output (summative), a mark?
4. If the judgmental rule yields a zone of uncertainty between any two preordained categories of judgment, A B C D E, what processes enable the teacher to assign a mark up or down? How and why do they work?
5. Do the identified cognitive processes form a pattern, schema or model of the marking process?
6. Do the identified teacher cognitive processes account for the five functions ascribed to marks by society in general?
7. Of the four research methods used in this investigation, is one superior for illuminating the marking process?

In addressing these questions, five elementary teachers from one Michigan school district served as research subjects. Data were gathered from each in the curriculum areas of language arts and mathematics following three marking periods across the school year. The methods of gathering and analyzing data

the range of values of α for which the function

$$f(x) = \frac{1}{x} \int_0^x t^{\alpha} dt$$

is continuous at $x=0$.

Let $f(x)$ be a function defined on $[0,1]$ such that

$$f(x) = \frac{1}{x} \int_0^x t^{\alpha} dt$$

included: process tracing, policy capturing, attributional theory and utility theory.

This chapter contains a description of the school setting along with the method of selecting teacher participants. The description is followed by an explanation of the procedures and methods used in data collection and analysis. The methods explanation is presented in the context of the judgment literature, with a rationale for the use of multimethod analysis of the self-report data which is basic to cognitive studies.

Research Setting

School District B, the site of this study is representative of the typical, suburban districts in Oakland County, Michigan. Its enrollment is declining, with a current pupil population of 14,500. These pupils are distributed across six secondary campuses and 21 elementary buildings. The pupils in District B come from a broad range of socioeconomic background, although ethnic mix is modest and racial mix minimal. Pupils in 10 of the 21 elementary buildings are served by Title I programs, indicating low socioeconomic status, while the majority of pupils in some buildings come from homes where the parents are professionals. Frequently these backgrounds are mixed in one building. Declining enrollment continues to cause mergings of these differing student populations.

Pupil performance in District B is average among Oakland County's 28 school districts. The district ranks in the middle of Oakland County's range on the Michigan Assessment Test. Performance on the California Basic Skills Test registers slightly above the national average. Pupil scores from the Differential Aptitude Test also support this average profile.

District B has a policy of building autonomy whereby principals and their staffs select their own pupil reporting system. Fourteen of the elementary schools report pupil progress, at the upper elementary levels, via traditional marks plus a checklist and comments. The remaining schools use checklists and written comments without marks. All schools have four marking periods and two parent/teacher conferences following the first and third markings.

Five principals from schools using traditional marks expressed an interest in the marking study, and said they would consult their teachers regarding participation in the investigation. (Experience beyond five years in the upper elementary classrooms was the only criterion of the researcher.) The first two principals contacted by the researcher, following the initial expression of interest by the five, yielded five volunteer teachers, three men and two women, from grades four, five and six. These five teachers became the subjects of the study.

The subjects were representative of the mode of teacher tenure in the district—none had less than 14 years of teaching experience. It is important to note that these subjects were not selected for being the "best" teachers. Instead they volunteered to give information to the investigator during free periods or when the principal substituted. Each teacher carried a typical class load ranging from 29 to 33 students.

Procedures

Data Collection

The structured interview was the primary source of data acquisition. Each of the five subjects was interviewed immediately following the first, second and final marking, of four periods. The interviews were taped and subsequently transcribed into protocols.

The in-depth interview is frequently criticized as a method of self-report which reflects known biases and justifications. In this study of marking, the strength of the interview is the potential to elicit personal opinions, including biases, knowledge and understandings related to one repetitive and official portion of the teacher's job. The potential is best realized by heeding expert advice on the interview.

The interview as a method is well discussed by Bussis and Chittenden (1976), Schatzman and Strauss (1973), Pelto and Pelto (1970). Each of these authors notes the enormous literature on interviewing formats and techniques, each stresses the importance of grounding interviews in other theoretical work, and each finds the information supplied by a key informant of a given cultural environment an essential element in constructing the meaning behind behavior.

The form of interview questions is important. "For many situations, fieldworkers should devise questions concerning concrete events, behaviors and possessions, instead of asking questions involving vague generalizations," says Pelto. Bussis and Chittenden reported that "questions phrased at a high level of generalization turned out to be answerable only by abstractions and generalities too vague to be revealing of personal constructs . . . and the type of question that more readily brought out personal constructs was one posed with concrete reference to materials, to classroom practices or to children's behaviors." Schatzman and Strauss state that the level of abstraction or concreteness is best found by becoming familiar with the culture, hence, "early interviews tend to prove less economical than later ones, mainly because the researcher has not yet fully determined precisely what information he needs Observation, brief questioning and casual conversation are so very important; they eventually provide a broad context for effective and economical interviewing."

Thus, the taped interviews in this study, conducted on-site, were based on previous insights into interview formats and focused on products of the teachers' own creation, such as, record books. This allowed for prediction, reflection and open-ended responses. The interview is presented in Appendix A.

In addition to the interview, data were collected from official marks, record books and a pupil sort. Marks of all students in each class were collected in language arts and mathematics although the teachers also marked in spelling, reading, social studies, science and art. The teachers were also asked to predict the marks for each student for the next marking period and give brief reasons why the mark was predicted to remain the same, go up or go down. The record book data were collected to cross-check teachers' verbal protocols.

Data Organization

The data collected were organized into six sections, a composite case and five individual teacher cases. The composite case includes a model of the teacher judgment processes during marking and subsections on rules, statistical analysis and protocol analysis. The five teacher cases, each of which also has a subsection on rules, statistical analysis and protocol analysis, follow the composite case.

Data Analysis

The analysis of data—marks, predicted marks, record books and pupil sort—was both qualitative and quantitative. Specific analysis of marks and predicted marks involved multiple regression analysis, Pearson correlations and frequency distributions. Transcribed interviews were coded verbatim and categorized in several ways: by the common attributional categories of ability, effort, task difficulty and home support; by elaboration of description; and by a decision tree.

Methods

Four research strategies from the field of judgment seemed especially congruent with the marking phases: process-tracing techniques to establish the validity of an overarching schema (taped interviews and content analysis of verbal protocols); policy-capturing techniques to analyze the record book system and combination rule throughout the year (multiple regression, Pearson and partial correlations and frequencies); utility analysis techniques to investigate teacher method of assessing risk related to classroom behavior (decision tree); and attributional techniques to investigate teacher method of assessing risk related to future student motivation to achieve (interview data related to record book analysis and prediction data).

Using a multimethod approach to explore teachers' grading processes, allowed the broadest description of the task. An integration of approaches sought to use the strengths of each method while minimizing the weaknesses by carefully distinguishing findings which were corroborated by several methodological perspectives and those which emerged in only one field of reference. In this manner, the study attempted to recreate teachers' understandings surrounding the judgment task and to relate the task to achievement and management in each teacher's unique classroom.

Process Tracing

The emphasis in the "process-tracing" method is on building a model or schema of the cognitive rules and heuristics used by clinicians. Analysis of verbal data from taped interviews or think-aloud sessions is the basic method of identifying implicit rules and heuristics. In contrast to the familiar linear regression technique of policy capturing which seeks to identify the judgment cues

which carry the most weight, process tracing identifies many cues which may have been used in a task but which do not necessarily receive the most weight (Einhorn, 1979). In the initial search to discover the underlying processes in the marking judgment, the researcher was interested in all cues and possible patterns which may serve as an adaptive mechanism between a very complex learning environment and a simple, required response (a mark).

Judgment during the marking process suggested two subjudgment phases, one which allowed multiple categories (five marks, A B C D E, with or without pluses and minuses) and a potential second one which allowed only a choice between any two categories. The decision between multiple categories often necessitates a quantitative mode of analysis while a choice between two, especially if one is considered a reward and the other a punishment, may dictate more qualitative approaches (Einhorn, 1979).

For the purpose of inquiry into a new cognitive arena, teacher marking, it is important to note that process tracing was used to establish a path in the wilderness, a pattern or schema. In many previous judgment studies, process tracing was used to indicate that a linear model was too simple when cues were used concurrently or as trade-offs. In such studies, cues were already specified and the method sought to assess how they combined, balanced or traded off. It is in this multiplicative sense that Einhorn describes process tracing as a more detailed method of analysis than linear regression when he critiques both (Einhorn, 1979). In this study, however, process tracing was much more general than linear regression, and it attempted to pattern decisions made over a longer time frame. The intended outcome of process tracing here was to build a general model. If process-tracing methods support a schema, then policy-capturing techniques and attributional-utility techniques could give a more detailed analysis of particular heuristics used at different phases of the task.

Verbal report data are frequently dismissed as unscientific and simply a form of introspection which is worthless for verification. Ericsson and Simon point out, however, that behaviorism and allied schools of thought have been schizophrenic about the status of verbalizations as data because the basic behavioral data in standard experimental paradigms are a signal "yes" or "no" or a choice between words which is essentially indistinguishable from verbal responses. But Ericsson and Simon also point out that researchers using verbal report data are remiss in reporting the details of how they collect and analyze verbal data and how it relates to even a rudimentary theory of cognitive processing. Ericsson and Simon are especially anxious to distinguish between concurrent verbalization which gains information from subjects while they are attending to it and retrospective verbalization where the subject is asked about cognitive processes which occurred at an earlier point in time (Ericsson & Simon, 1980).

These concerns parallel the interest of researchers who have constructively criticized the in-depth interview. The strength of the interview is its potential to elicit personal opinions, knowledge and understandings. Within this study the interview has been detailed under Data Collection. In an attempt to strengthen insights gained from the interviews, multimethods of analysis have been applied and the interview process has been grounded in official records, teacher marks.

Policy Capturing

The most frequently used method of studying judgment processes is policy capturing (Shulman & Elstein, 1975; Slovic & Lichtenstein, 1971). Beginning with a linear model, this method attempts to reproduce the inferential responses of a particular judge as he weighs and combines important factors (cues) in the task environment. Policy capturing concentrates on identifying these cues, defining

their relationship to each other, and estimating their relationship to a final judgment for purposes of prediction and control.

Policy-capturing studies generally discuss alternative methods of cue specification. Five approaches are delineated and evaluated in an article by Clark, Yinger and Wildfong (1978): (1) logical specification, (2) expert opinion, (3) prespecification and narrowing of a large number of potential cues, (4) allowing cues to emerge during a judgment task and (5) participant observation in a naturalistic setting. The expert opinion method consults experts to determine their nominations for important cues in a specific judgment task. This approach has been widely used in studies of pathologists, physicians and stockbrokers (Shulman & Elstein, 1975). But the approach is criticized because cues which may be nominated may not be the ones actually used in practice and/or because experts may weight cues very differently.

Policy capturing's reliance on linear regression has also been criticized as much too simple a depiction of human judgment with its highly contingent methods. Einhorn has repeatedly responded to this "simplistic" label by pointing out that the mathematical representation of a model should not be confused with the process which it represents. He elaborates that the essence of linear models is that they imply trade-offs thereby "not only acknowledging the existence of conflict but resolving it through compromise . . . linear models of judgment imply a sophisticated and complex cognitive strategy" (Einhorn, 1979).

Policy capturing lends itself to a study of the processes involved in teacher judgment during marking due to the specific nature of the task environment. Cue specification in the marking task relies on the teacher as expert with continuous corroboration from official records in the grade book. Verification of judgments can come through retrospective generalizations which rely on past action that is evidenced in public record. Stenhouse proposes "the

ideal that no qualitatively-based theorizing in education should be regarded as acceptable unless its argument stands or falls on the interpretation of accessible and well-cited sources so that the interpretation offered can be critically examined" (Stenhouse, 1978).

A routine check of the record book quickly established how the cues were combined to arrive at the judgment (mark). A strategy of recording and predicting grades over a year's time (three marking periods) also provided statistical evidence of trends across time, correlations between marks and predictions and identification of anchoring or recency effects.

When policy-capturing studies attempt to infer causes for the choices of a judge or weights assigned to cues used in a judgment, they must guard against two common errors. Kahneman and Tversky (1973) have named these the "availability" and "representative" biases. The first infers cause from available evidence, apparent frequency or accessibility in present memory. Obviously biases in the judge's exposure, level of attention and storage can arise. The representative bias predicts future choices by reflecting on the degree to which the specified outcome represents its origins or resembles a like case. In drawing inferences, the decision heuristic of which class or category an event represents, is an essential tool which must be understood to avoid bias such as the familiar "Gambler's Fallacy" which leads one to conclude after observing a long run of red on a roulette wheel, that black is now due because it has been underrepresented in a chance process up to now.

Preliminary examination of teachers record books indicate an intent on the part of teachers to "be fair" and to eliminate the common biases referred to by Kahneman and Tversky. (Sample record book in Chapter IV.) The grid of the record book becomes an inferential teacher tool to arrive at a baseline figure of performance within a class population (vertical column) balanced against a

baseline figure for individual performance (horizontal). The baseline for whole class performance is related to the assignment which is considered to be indicative of similar (representative, normed) grade level performances across the nation according to the textbook publishers and related to the collected performances (consensus) of that particular class by collected samples of work graded on a point basis. The baseline data for a given student are the collection of numerous work samples of the same student over time. The record book, therefore, is used as a scientific data base.

The policy-capturing phase of this study analyzed this data base by multiple regression methods in order to capture some potential underlying policies of the teacher. The information obtained lead to an understanding of "what" work samples a teacher accumulates and "how" the specific marking judgment is made.

Utility Analysis

Although policy capturing appears well suited to the first phase of marking, it does not shed sufficient light upon the process if the combination rule does not yield a perfect preordained category such as A B C D E. What process operates when a student's work falls into a zone of uncertainty midway between two marks? How does the teacher decide to go to the higher or lower mark?

Utility analysis is a systematic approach to decision making under conditions of uncertainty and risk. It forces the decision-maker to separate the logical structure of the decision problem into its component parts to be analyzed individually and recombined systematically (Weinstein, Fineberg, Elstein, Frazier, Newhauser, & Neutra, 1980). Alternative actions must be clearly described and the question of the need for additional information at each branch of the decision tree is recurring. Alternative outcomes, assessment of values and possible trade-offs are crucial to decision analysis and involve utility decisions.

The decision tree is the fundamental analytic tool for utility analysis. It requires the decision-maker to identify alternative actions which might be taken at different times and to obtain information at these times. An optimal course of action is desired.

The building blocks of the decision tree are:

1. Choice nodes, at which one or two alternative actions may be explored.
2. Chance nodes, at which the status of the student is revealed, test information becomes available or other outside information.
3. Outcomes, which describe what happens to a student along each path of events in terms of attributes held to be of value (i.e., classroom cooperation and concentration on task).

"By convention, a decision tree is built from left to right, with choice nodes represented by squares and chance nodes by circles and with outcomes specified at the right hand 'tips' of the tree" (Weinstein, Fineberg, et al., p. 16). A decision tree for the marking process may be found in Chapter IV, Figure 4.4.

The reasoning process behind utility analysis can be summarized in four major stages:

1. Data acquisition, in which information is obtained by the teacher using a variety of methods (tests, homework, oral participation in class, observed behavior during class sessions).
2. Hypothesis generation, in which alternative problem formulations are retrieved from memory.
3. Data interpretations, in which the data are interpreted in light of the alternative hypotheses under consideration.
4. Hypothesis evaluation, in which the data are used to determine if one of the diagnostic hypothesis already generated can be confirmed. If not, the problem must be recycled: New hypotheses generated, and additional data

are collected until one of the hypotheses is verified (Weinstein, Fineberg, et al.).

Weinstein, et al. point out that two modes of clinical inference are observed in the process; one is diagnostic and the other therapeutic. In the case of teacher marking judgments, both are concerned with sustaining or increasing a student's motivation and cooperation as indicated by work production, class participation and lack of problem behavior.

The decision-analysis process appears well suited to investigate a teacher's choice of giving a higher or lower grade, and each marking period gives the teacher a new chance to assess the hypothesis by examining the record book and reflecting on class behavior.

Attribution Theory

Attribution theory is concerned with how a person infers or attributes cause to an event and what happens once he does. Attributional research goes beyond a description of what factors are in evidence during judgment to an explanation of why and how the factors are combined into a decision. Two major theoretical approaches to attribution are appropriate to the marking decision; the covariation principle of Kelley and the dimensional approach of Weiner.

Kelley (1971) proposed that people determine the cause of behavior by examining the covariation between the effects and possible causes of the behavior and then attributing the effect to the cause or causes with which it covaries. Covariation of an effect over time is called "consistency." The extent to which an effect is localized across situations is called "distinctiveness" and the extent to which an effect is found across people is called "consensus."

The application of Kelley's covariation principle to teachers' marking judgments emphasizes the extent to which a student's effort and ability are judged to be within his own responsibility and control. It offers specifications of data about a student's effort (work samples over time) and ability (test results) and contingency factors (parental support/pressure, attendance/health) which can lead to the assignment of high or low marks. The effect of consistency is that the more often a student turns in work, the more likely he is to be judged responsible for learning the work. The effect of distinctiveness is that the more the student misses work assignments only in restricted circumstances, the less he may be held responsible for the failure (i.e., when ill or when evidence of parental problems is overwhelming such as death, divorce or abuse). The effect of consensus is that the absence of work samples and reasonable test results when associated with few students in a given class is judged the responsibility of the student more than when associated with many students in the particular class which may indicate external factors such as task difficulty are involved.

The role of consensus or base-rate information is controversial in many fields. Its usefulness in attributing cause has been detailed in the research of Kahneman and Tversky (1973) and Nisbett and Ross (1980). In the marking process, it is important to note the consensus of one classroom may differ from another (especially if there is a tracking system) and from one school environment to another (especially if there are large socioeconomic differences). This area is well discussed in the teacher expectation literature and will not be delineated further here except to note that for an individual teacher, the consensus of a given elementary, self-contained classroom is primarily limited to the boundary of that classroom system or culture.

Weiner's (1979) dimensional theoretical approach, developing from concepts of Heider (1958), and Kelley (1967), provides evidence that attribution

judgments for success and failure are divisible into three dimensions: stability, locus and control. Stability considers whether the causes of behavior are stable or variable over time and contrasts Heider's notions of relatively fixed characteristics such as ability and typical effort with fluctuating factors such as immediate effort, attention and mood. Locus considers whether causes are within (internal) or outside (external) the person, and from the perspective of achievement, internal causes may include ability, effort, maturity, etc., while external causes may include classroom environment, task and family. Control considers the degree to which the individual is responsible for the present event, hence, the degree of control of future change.

Weiner proposes that these three dimensions create a meaningful grid on which to assess the myriad causes of achievement events. Furthermore, although numerous perceived causes of success and failure are discussed in the literature (Cooper & Burger, 1978), there is rather a small list from which the main causes are repeatedly selected. Within this limited list, ability and effort appear to be the most salient causes.

Weiner contends that each of the three dimensions of causality has a primary psychological function or a linkage. The primary relation of the stability dimension is to the magnitude of expectancy change following success or failure. The locus dimension has implications for self-esteem and the perceived control dimension relates to helping, evaluating and liking by others.

Stability. Attributional research in regard to stability has remained remarkably consistent since 1971. "Empirical evidence . . . has proven definitively that causal ascriptions for past performance are an important determinant of goal expectancies" (Weiner, 1979, p. 9). Failure ascribed to low ability or to task difficulty decreases the expectation of future success more than failure ascribed to bad luck, mood or a lack of immediate effort (Weiner et al., 1976). In 1976

Weiner et al, carried out an investigation which clearly discriminated between the dimensions of stability and locus, and proved that expectancy changes are related to stability and are not associated with locus as was previously postulated in substantial competing literature. Weiner concluded that "the literature associating stability with expectancy change is unequivocal, and the findings generalize outside of the laboratory as well as beyond achievement domain" (Weiner, 1979, p. 10).

A comparison of Kelley's covariation principle and Weiner's dimensions points to considerable overlap in categories such as consistency and stability, distinctiveness and locus, however, Kelley's theory is primarily historical record while Weiner's is primarily concerned with future control. Together they shed great insight on the cognitive process used in the teacher's marking task.

Locus. The concept of locus is not as firmly established in research. Weiner, himself, has corrected an earlier position which presumed an invariant positive relation between internality and maximum emotional reaction (Weiner, 1971, 1977 and 1978). A series of studies was initiated to determine the relation between attribution and affect in which subjects responded to a series of scenarios depicting success or failure experiences. About 100 affects for success and 150 for failure were provided with a rating scale for intensity. The findings indicated a two-phase response, the immediate and intense affect was related to outcome regardless of "why." That is, success resulted in obvious happiness, and failure revealed displeasure and upset. But attribution linkages were also present. Given success, the linkages were: ability - competence + confidence; typical effort - relaxation; immediate effort - activation; others - gratitude; personality - pride; and luck - surprise, relief and guilt. For failure, the attribution linkages generally were: ability - incompetence; effort - guilt and shame; personality - resignation; others - anger and aggression; and luck - surprise. In the long run, central self-

esteem emotions that facilitate or impede subsequent achievement may be more effected by the attributional linkages than by the immediate affect of feeling good or bad.

Control. The control dimension concerns inferences about the intentions of others rather than causes. This dimension is perhaps more meaningfully related to the teacher's role of helping, evaluating and liking (sentiment). The majority of investigations into helping behavior concluded that help is more likely when the perceived cause of the need is an environmental barrier as opposed to being internal to the person desirous of aid (Berkowitz, 1969; Piliavin, Rodin, & Piliavin, 1969). Piliavin found that when a failure is perceived as controllable, then help is withheld under the assumption that the person should help himself. Hence a drunk person is less likely to receive aid than an ill person since illness is considered uncontrollable.

One experiment on helping concerned altruism in the classroom--lending class notes to an unknown classmate (Barnes, Ickes, & Kidd, 1977; Weiner, 1979). In this experiment, two causal themes for a student's lack of notes were contrasted in eight possible combinations. In one theme the student always (stable) or sometimes (unstable) did not take notes because of something about himself (internal) or something about the professor (external). The student was unable to take good notes (uncontrollable) or did not try (controllable). Following each causal statement, subjects on a 10-point scale, rating from "definitely would lend notes" to "definitely would not lend notes." The results indicated that help is reasonably extended in all combinations except when the cause is internal and controllable such as if the student did not try to take notes or in the second theme if the student could have avoided being absent.

Data from investigations into evaluation "conclusively demonstrated that effort is of greater importance than ability in determining reward and

punishment. High effort was rewarded more than high ability given success, and lack of effort was punished more than lack of ability given failure" (Weiner, 1979). Weiner explains this discrepancy between ability and effort by referring to moral feelings of "ought to do" and by the feeling that reward and punishment could actually change behavior in the future. "That is, there is a pervasive influence of perceived controllability or personal responsibility . . . in achievement related contexts, including how students are graded" (Weiner, 1979, p. 17).

Attribution analysis involves a schematic model of the attributional process which is clarified by further investigation of a specific set of cues and related by the person to a possible set of outcomes. "In a typical study, the subject is provided with a number of informational cues in all possible combinations and asked to rate why each event described by a particular cue combination may have occurred. ANOVA models have frequently been used for data analysis, and interaction effects have been equated with configularity in cue usage. Although these studies have used a wide variety of cues, situations, and attribution rating scales, results have been surprisingly consistent" (Carroll, Payne, Frieze, 1976).

The attribution and decision-making framework of the current study attempts to elaborate one phase of the overall marking process, the phase wherein the teacher assesses the risk of obtaining more effort (homework and tests) and more achievement (tests) by giving a higher or lower mark at a marking period. Through interviews which call for prediction of future marks, accompanied by the factors (cues) influential to such a prediction, and the eventual outcome of the prediction (the grade), this investigator attempted to flesh out a more complete schema of the marking process.

Summary

This investigation into teacher judgment during the marking process attempted to answer seven research questions. Data were gathered through in-depth interviews with five elementary teachers from one Michigan school district across one school year. Interview data were grounded in the teachers' record books and the official marks of 152 students. A multimethod approach was used in order to provide both quantitative and qualitative analysis. The four established judgment approaches (process tracing, policy capturing, utility analysis and attribution theory) allowed for cross-checking and corroboration of evidence thereby lessening the risks of total reliance on self-report data. The evidence gathered was organized into five teacher cases and one composite model of the teacher marking judgment.

CHAPTER IV

FINDINGS

Introduction

The major purpose of this study was to create an understanding of the judgment processes which engage teachers during marking across a school year. Four established judgment methods were used to uncover the process; process tracing, policy capturing, attribution and utility analysis. The basic data was collected through taped and transcribed interviews. A unique aspect of this multimethod analysis was that the interviews were grounded in official public records; the teacher record book and summative marks on report cards. All interviews were done with the record book at hand. The public aspect of marks added an extra dimension of objectivity to an otherwise subjective process.

In this chapter, the findings are organized to display the data which teachers considered during the marking task, to make inferences from the data and to ascertain whether or not (1) the data indicates a coherent and meaningful process which suggests a model and (2) whether or not the data answers the major research questions which guided the study (page 11). The findings are presented as six cases beginning with a composite case which serves as an organizer followed by five teacher cases. Each case is further divided into four sections; marking rules, statistical analysis, verbal analysis and a summary. An integrated summary concludes the chapter.

The composite case format was chosen to serve as an organizer, setting a pattern for the individual cases. Basic data are displayed in each teacher case and inferences are made. In order to avoid lengthy repetition of common discussion points within each teacher case, the inferences are more extensive

within the composite case. Discussion within each teacher case will then refer back to the composite case, noting points of difference. This format results in a detailed composite case followed by more concise individual cases, although it must be kept in mind that the data were originally collected on an individual basis and later combined to create the composite model.

Composite Case

The composite case depicts factors which the five teachers have in common. These commonalities were derived through computation of the marking data of all five teachers using multiple regression, correlations, and frequencies (N.I.E., SPSS, 1980) and through content analysis which distilled common rules, categorized and coded attributes and utilities and identified key descriptions from the interviews.

Rules

The process tracing-phase identified two sets of rules which guide the marking judgment: procedural and contingency. The two types of rules dealt with different aspects of the judgment process. From the point of view of information processing, procedural rules were concerned with selection and simplification. These rules set up a linear, routine record book system, selected tasks for inclusion and accounted for academic standards and precision measurement for marks on tasks. Procedural rules were product and time based and lent themselves to statistical analysis.

Contingency rules were those which determined judgment in uncertainty and exception. From the point of view of information processing, contingency rules were concerned with inferential processes which go beyond the data. These

rules were essentially involved with factors which promoted (1) stable individual task completion over a year's time, and (2) a stable classroom environment for on-task behavior or class flow over a school year.

Contingency rules involved teachers in an assessment of motivational factors for each student, including ability, effort, home support, classroom behavior and task difficulty. Hence these rules were motivation and behavior related and lent themselves to verbal analysis. The rules, distilled from transcribed interviews, highlighted these two major aspects; one of routine judgment procedure and one of contingency judgment strategies.

Procedural rules:

1. Teachers assumed that completed tasks resulted in learning (implicit, not stated).
2. Teachers assigned tasks and gathered marking data regularly in a record book.
3. Teachers accounted for task completion at a given level of difficulty with a check system and for task completion at a given standard of mastery by a mark.
4. Teachers gathered marks from a sufficient variety of tasks (tests, written projects, exercises) to satisfy their criteria for validity. None had less than six formative marks. Four had more than ten.
5. Teachers had individual theories about weighting some tasks (tests vs. homework) more heavily than others.
6. Teachers had individual systems for transforming points representing standard criteria on a written paper into A B C marks.
7. Teachers had a combination rule for transforming formative marks into summary marks. They added all task marks across and divided by the total number of assigned tasks (arithmetic mean). This was corroborated by an analysis of each record book in math and language arts.

Contingency rules:

1. Teachers ranked effort as related to ability as a prime criteria for going up or down. Effort was judged by regular work and extra work (record books and attribution chart).
2. Teachers had strategies to apply if the work fell midway between two marks.
3. Teachers had individual strategies for marks which fell below C (frequencies and quotations).

Procedural rules resulted in a record book system which operated as a statistical tool to help overcome many of the common errors of human judgment, which were discussed in the preceding chapter. An analysis of the record book showed the teachers' intent to account for a base rate of work for the nation, (the assignments adjusted to grade level on nationally normed verbal information, i.e., textbooks), for the classroom (the vertical column of any given assignment) and for the individual student (the horizontal row). Hence teacher record books were inferential tools that depicted student achievement in comparison to individual ability, class (group) and nation.

Initially the teacher used only the record book to compute the mark into a preordained category of A B C D E. However, when the work fell into a zone of uncertainty between two grades or when it fell into the D or E category, contingency rules were put into operation. A statistical analysis of the marks (152 students) which resulted from the procedural rules follows.

Statistical Analysis

The statistical methods involved multiple regression analysis, Pearson and partial correlations, frequency counts and cross tabulations. Marking data was entered in the computer under language arts and mathematics. The following symbols were used in all charts contained in this technical report. L_1 representing

the first mark in language, L_2 the prediction of the second language mark, L_3 second mark in language, L_4 the prediction of the final language mark, L_5 final mark in language; M_1 representing the first mark in math, M_2 the prediction of the second math mark, M_3 second mark in math, M_4 the prediction of the final math mark, M_5 final mark in math.

For computation purposes, the summative marks, the predicted marks and the final marks in both language arts and mathematics were arithmetically valued and entered in the computer as follows:

A+ = 13	B+ = 10	C+ = 7	D+ = 4	E = 1
A = 12	B = 9	C = 6	D = 3	Incomplete = 0
A- = 11	B- = 8	C- = 5	D- = 2	

These values were used to derive all statistical factors which are found within the figures and tables of this study. Their role is particularly told in the composite teacher policy model, Figure 4.2.

Across the year. Multiple regression is a general statistical technique for analyzing the relationship between a dependent variable and a set of predictor variables. In this case the researcher was interested in the extent to which the final summative mark in a subject is a function of preceding marks within the same year. The most important use of the regression technique is to describe the best linear prediction equation. Using multiple regression analysis, composite teacher marking policies (equations) for language and mathematics were captured:

$$L_5 \text{ (final language mark)} = .42L_3 + .30L_4 + .16L_2 + .15L_1 - \text{Constant}$$

$$M_5 \text{ (final math mark)} = .39M_4 + .34M_3 + .23M_1 - \text{Constant}$$

There is some inconsistency between the policy equations. In the language policy, the greatest predictor of the final mark is the second mark (Beta weight = .42). The second greatest predictor is the final prediction (Beta = .30) and the poorest predictor is the first mark (Beta = .15). In the math equation the final prediction is the greatest predictor of the final mark (Beta = .39) and the second mark (.34) is the second greatest predictor.

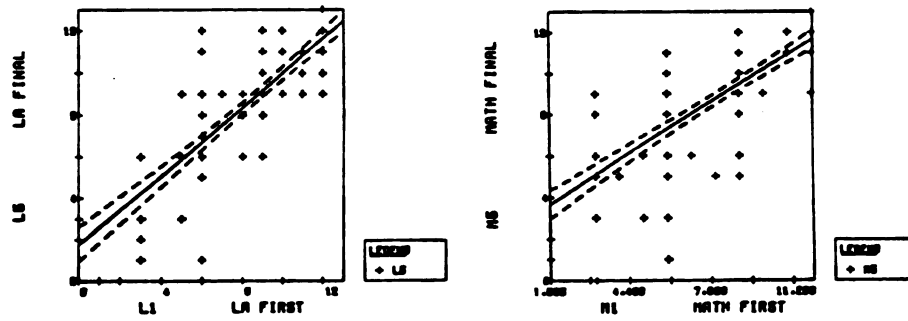
All composite regressions were plotted. The basic purpose of a plot is to give a pictorial representation of the relationship between any two variables and to indicate the confidence interval of the regression line. The relationship between the first and last marks in language and math are depicted here. All other regression plots are in Appendix D.

Table 4.1

Relationship Between the Final Mark and Predictor Marks

Teacher N = 5					Student N = 152				
Dependent Variable = L_5 Final Language Mark					Dependent Variable = M_5 Final Math Mark				
Language Arts Variables		B	F	Sign.	Math Variables		B	F	Sign.
Final Prediction	L_4	.30	6.	.016	Final Prediction	M_4	.39	16.	.000
Second Marking	L_3	.42	18.	.000	Second Marking	M_3	.35	16.	.000
Second Prediction	L_2	.16	4.	.063	Second Prediction	M_2	—	—	.087
First Marking	L_1	.15	3.48	.064	First Marking	M_1	.23	11.	.001
Note. Overall F = 171 Sign. .000 Multiple R = .91 R Square = .82 Standard Deviation = 1.26					Note. Overall F = 168 Sign. .000 Multiple R = .91 R Square = .83 Standard Deviation = 1.27				

Note. Multiple regression analysis was used with the marks of 152 upper-elementary students to describe the best linear prediction equation. It is important to notice that not all Beta weights are significant in the equation.



Note. The relationship between the first and final mark is graphically presented. Of importance is the depiction of the confidence interval surrounding the regression line.

Figure 4.1. Plotted relationship between first and final marks.

The possible inference that the final mark is largely a function of the second mark and the final prediction, is supported by the Pearson correlations. Pearson correlations are statistical techniques which provide a single number to summarize the relationship between two or more variables. In the case of marks, the researcher was interested in the relationship between the final mark and preceding marks and predictions. The correlational charts included here indicate that the final composite language and math mark is more closely correlated to the actual second mark (.88) and the final predicted mark (.88) than to the first mark (.82).

Table 4.2
Correlations Between Marks
and Predicted Marks Across a Year

Teacher N = 5

Student N = 152

Language Arts Variables

L₁

L₂

L₃

L₄

L₅

First Marking

L₁

.89

.79

.83

.82

Second Prediction

L₂

.89

.78

.83

.81

Second Marking

L₃

.79

.78

.93

.88

Final Prediction

L₄

.83

.83

.93

.88

Final Marking

L₅

.82

.81

.88

.88

Mathematics Variables

M₁

M₂

M₃

M₄

M₅

First Marking

M₁

.88

.78

.78

.81

Second Prediction

M₂

.88

.78

.81

.78

Second Marking

M₃

.78

.78

.91

.88

Final Prediction

M₄

.78

.81

.91

.88

Final Marking

M₅

.81

.78

.88

.88

P = .001

P = .000

P<.001

P<.001

Note. The Pearson correlations were derived from the summative marks (first, second and final) of elementary teachers across the school year.

Before drawing final inferences from the regression and correlational analysis, the policy equation must be adjusted by the evidence derived from partial correlations. Partial correlations allow the researcher to describe the relationship between two variables while adjusting for the effects of one or more additional variables. In this case the researcher was interested in double-checking the relationship between the predicted final mark (L₄ and M₄) and the actual final mark (L₅ and M₅) while controlling for other previous marks. Specifically when controlling for L₁ L₂ L₃, the correlation between the final prediction and the final mark is reduced from .88 to .20, a reduction which is also present in mathematics. The relationship between the prediction of the second mark and the actual mark (L₂ L₃), when controlling for the first mark (L₁) equals .26 and in math it equals .31.

Table 4.3
Controlled Relationship Between
the Final Prediction and the Final Mark

Teacher N = 5

Student N = 152

Language Arts	Controlling for:	Coefficient	P
$r(L_4, L_5)$	L_1	.64	.001
	L_2	.65	.001
	L_3	.37	.001
	L_1, L_2	.59	.001
	L_1, L_3	.24	.002
	L_2, L_3	.22	.004
	L_1, L_2, L_3	.20	.008

$r(L_2, L_3)$	L_1	.26	.001
---------------	-------	-----	------

Note. L_1 = First Actual Mark
 L_2 = Second Prediction
 L_3 = Second Actual Mark
 L_4 = Final Prediction
 L_5 = Final Mark

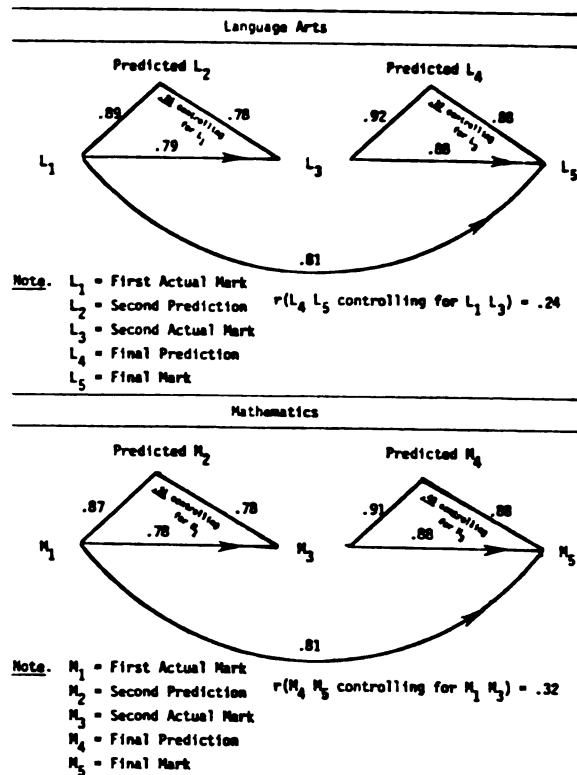
Mathematics	Controlling for:	Coefficient	P
$r(M_4, M_5)$	M_1	.67	.001
	M_2	.67	.001
	M_3	.40	.001
	M_1, M_2	.64	.001
	M_1, M_3	.32	.001
	M_2, M_3	.31	.001
	M_1, M_2, M_3	.31	.001

$r(M_2, M_3)$	M_1	.31	.001
---------------	-------	-----	------

Note. M_1 = First Actual Mark
 M_2 = Second Prediction
 M_3 = Second Actual Mark
 M_4 = Final Prediction
 M_5 = Final Mark

Note. Partial correlations were derived from the marks of 152 upper elementary students across the year.

Through statistical analysis, the following conclusions appear. All marks are highly correlated and significant ($P = .001$). The final mark is most highly correlated to the second mark which is primarily correlated to the first. Correlations between actual marks are greater than those between predictions and marks, although it is obvious that predictions become more closely correlated with actual marks as the year progresses. The following judgment model emerges:

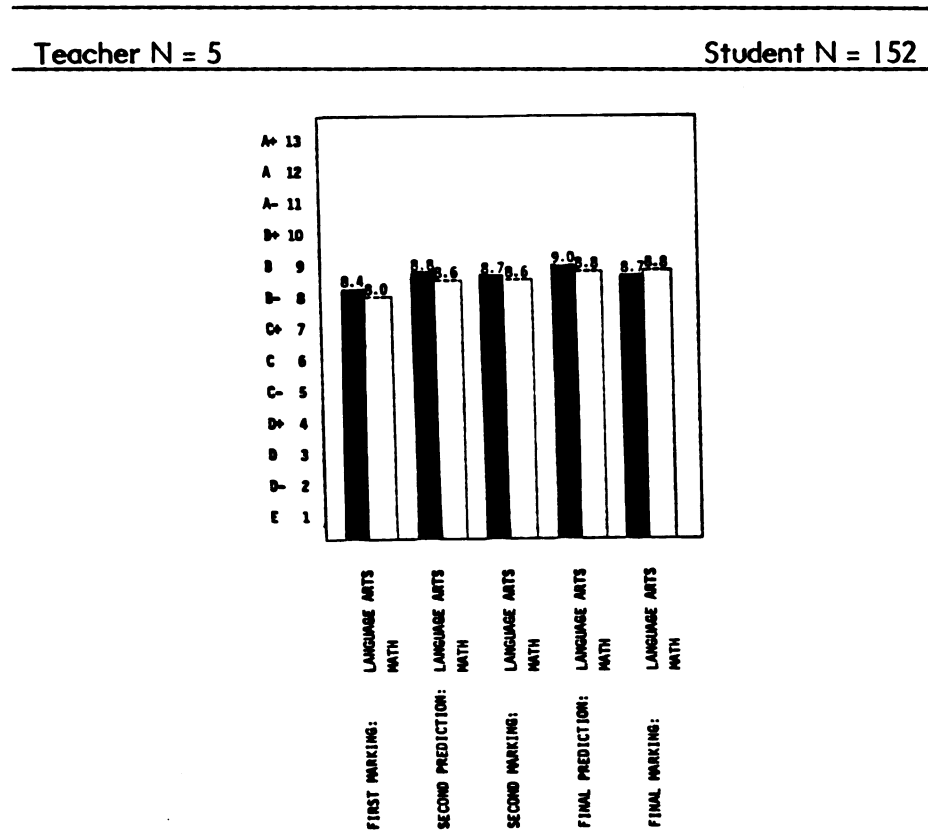


Note. This judgment policy or model was derived from the statistical analysis portion of the judgment study (multiple regression, Pearson correlation and partial correlations) and is based on the summative and predicted marks of 152 students across a school year.

Figure 4.2. Composite marking policy (with predictions).

The judgment model is corroborated by the pattern of the bargraph (frequencies) which illustrates that the average marks across the year are generally slightly lower than predictions.

Table 4.4
Composite Pattern of Marking Averages Across a Year



Note. Bargraph derived from teachers' average marks and predictions over a year's period.

At this point in the analysis, using statistical techniques, three competing hypotheses are possible; a recency effect, a primacy effect and an averaging. The original regression equation suggested that the marking policy was most heavily influenced by a recent mark or prediction (a recency effect). This hypothesis represents a phenomenon noticed in many theories of error in human judgment in which thinking is determined by the most recently available information in the memory structure. However, this conclusion is drawn into question by the partial

correlations. There is evidence in the partial correlations for the hypothesis that the final mark is "anchored" in the first mark. The anchoring heuristic suggests that teachers determine a student's marks early in the year, and adjustments thereafter are biased toward the initial values. In the judgment literature, this anchoring phenomenon is a typical error of laymen (Nisbett & Ross, 1981; Kahneman & Tversky, 1974), and it was found to be typical of teacher expectancy/behavior patterns in repeated classroom interaction studies (Brophy & Good, 1970). Therefore, the inference that the last mark is anchored in the first mark is not only present in the data but established in the literature. These two hypotheses offer explanations at both extremes of the spectrum.

A third logical inference is possible and is not eliminated by the data. The final mark may not be anchored in the first mark or recently determined by the latest mark, but in fact, all three may be grounded in the teachers' record books. This inference is not apparent in either the regression equation or the partial correlations. However, regression and correlational analysis rely on arithmetic abstractions which may avoid some important subtleties of the marking judgment. A more detailed examination of the record book and the frequencies of marks within each marking period was undertaken.

Within marking periods An analysis of the teacher record books indicated that each marking period stood on its own formative marks. Teachers regularly assigned tasks and indicated completion in the record book. They stated that they followed a combination rule which involves an arithmetic mean. That is, either by hand calculation or calculator, teachers added all formative task marks and divided by the total number of marks. Departures from this rule resulted under two circumstances: when the formative work fell between two preordained categories and when the work fell below C. Corroboration with records indicated that this was generally followed. A formal statistical analysis of the formative

marks within the teacher record book was not undertaken. Instead, the data from the interview were corroborated by collecting the record books, photocopying all language and math marks, and cross-checking.

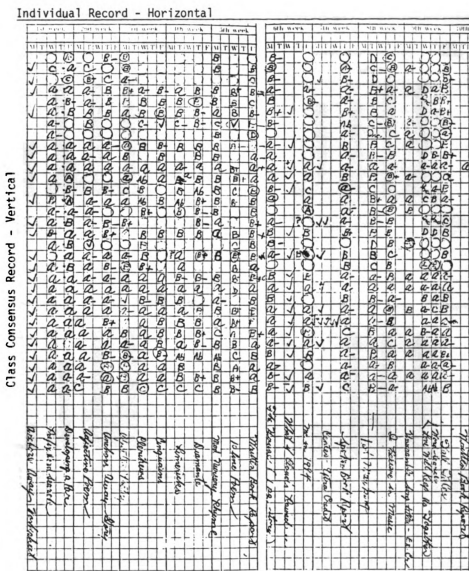
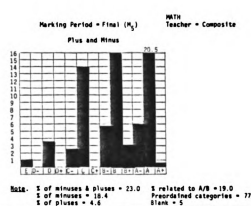
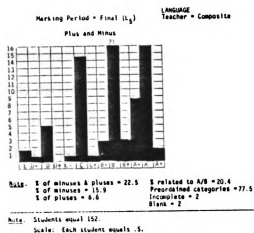
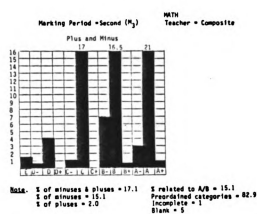
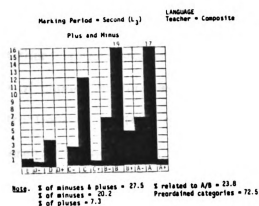
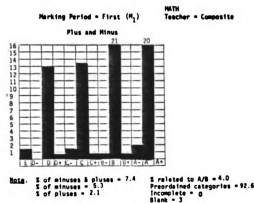
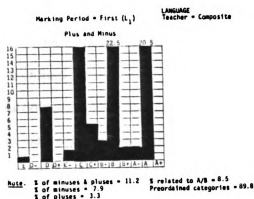


Figure 4.3. Sample record book account.

Table 4.5
Distribution of Marks Across
Three Marking Periods - Language and Math



An analysis of frequencies of marks, including pluses and minuses, indicated that there are patterns within marking periods. Before discussing inferences, it must be noted that two teachers did not give minuses or pluses on report cards; however, they did use them in the record book or they had a schema for extra effort (see Teacher Three and Teacher Four).

The majority of marks within each marking were assigned to the preordained categories with only a minority invoking contingency rules. Preordained categories predominated heavily in the first marking. Contingency rules operated more regularly after the first marking with minuses given more often than pluses. Minuses and pluses were primarily used as a contingency measure to qualify As and Bs. During the first language marking, only 2.7% were C or below; second marking only 3.7%; final marking only 2.1%. Only one D+ was given in language and math combined with two Ds. The inference may be drawn that there is more subtlety involved with marking at the upper level and that the rate of change in minuses and pluses does not show up clearly in the arithmetic abstractions of multiple regression and correlations. Contingency rules operate differently below C than above C, indicating that Ds and Es carry an altogether different connotation to teachers than marks which are average and above.

The greatest evidence for the influence of the formative tasks of a marking period in determining the summative mark was the official existence of the record book, the listing and marking of tasks completed, and the integrity of the math calculations. This was further supported by the fluctuations between marking periods.

It became evident that each marking period was self-contained despite potential hypotheses to the contrary. From the policy-capturing perspective, the marking judgment in the majority of cases was procedural and linear, albeit representative of only one phase of the process. Although the record book analysis

lends more detail to the overall understanding of teachers' judgment processes, neither the summative nor formative levels of analysis reveal the judgment factors operating within the contingency rules nor answers the questions why teachers do what they do. Other methods of inquiry were necessary to expose these processes.

Verbal Analysis

Verbal analysis of protocols, like the statistical analysis was put into perspective by the marking rules which emerged through process tracing. This part of the study was particularly concerned with identifying the judgment factors underlying the contingency rules.

Contingency rules are related to uncertain zones midway between marks and to cases of failure or near failure. Exposing the judgment cues involved various methods of establishing and categorizing teacher concerns. The interview process not only recorded marks of 152 students, but asked teachers to predict the next marking and to discuss the factors which influenced the prediction.

The common attributional categories were discussed in some detail in Chapter III. They are ability, effort, task difficulty and luck. In coding verbatim responses, a category of miscellaneous existed for one-time events. Early in the process, luck was replaced by an emergent home support category, and miscellaneous was replaced by an emergent category on class behavior and physical maturity. The later is more aligned with utility and maintenance of class flow or on-task behavior. Teacher statements were counted and turned into percentages. A copy of the coding device may found in Appendix C. The categories follow in Table 4.6.

0 0 100 000 1000 10000 100000 1000000 10000000 100000000

Table 4.6

Teacher Attribution-Utility Categories Upper Elementary Students

ABILITY (ACHIEVEMENT)

CONCEPT OF AVERAGE; ABOVE,
GOOD, MUCH BELOW, LOW, BELOW
GRADE LEVEL.
CONCEPT OF BRIGHT; VERY BRIGHT,
ABSORBENTLY TOP-NOTCH STUDENT,
BRIGHTEST KID IN CLASS, SLOW,
CONCEPT OF ACHIEVER; OVER/UNDER,
HIGH/LOW.
CONCEPT OF GRADES; A, B, C STUDENT
B-C, STRAIGHT-A STUDENT.
TEST FREEZE.
SPECIAL EDUCATION STUDENT.

TOTAL ABILITY.

EFFORT (MOTIVATION)

IN CLASS:
ATTENDING, CONCENTRATING.
WASTING TIME, LAZINESS, LACK
OF DISCIPLINE.
SPEED, CARELESSNESS, FINISHES
IN FIVE MINUTES.
TOTAL.

OUT OF CLASS:
CONSCIENTIOUS.
HAS POOR STUDY/WORK HABITS.
DOES EXTRA WORK, MORE THAN
IS ASKED FOR.
MAKES UP ALL ASSIGNMENTS.
WORKS AHEAD.
TOTAL.

COMPREHENSIVE:
OVERACHIEVING, REALLY TRYING.
UNDERACHIEVING, UNSTABLE EFFORT.
COMPETITIVE, KEEPS UP WITH FRIENDS.
"STIMULATING HIM TO DO ANYTHING
IS ALMOST A ONE-TO-ONE BASIS."
DETERMINED TO GET ALL A'S.
CAN'T GET HIS ACT TOGETHER.
NO MOTIVATION.
NOT MUCH ENTHUSIASM.
NEEDS TO BE PRODDED CONSTANTLY.
VERY DISORGANIZED.
TOTAL.

TOTAL EFFORT.

TASK DIFFICULTY

SKILLS
MULTIPLICATION NOT MASTERED.
DIVISION IS OFTEN DIFFICULT.
PROBLEM EXPRESSING IDEAS IN
WRITING.
STUDENTS SPEAK WELL SO WE'RE
WORKING ON WRITING.
ADDITION AND SUBTRACTION NOT
MASTERED.
MAY DIP AS CONCEPTS BECOME MORE
DIFFICULT.
TOTAL.

TEXT BOOK LEVEL
READING ABOVE GRADE LEVEL.
READING AT GRADE LEVEL.
READING A COUPLE OF GRADES
BELOW LEVEL.
SOCIAL STUDIES BOOK IS DIFFICULT.
SOCIAL STUDIES TESTS ARE HARD.
TOTAL.

GENERAL
LEARNING DISABLED.
CHILD IS BEING TESTED.
HAS A HARD TIME LEARNING MY GOALS.
TROUBLE CONCENTRATING.
BETTER IN LANGUAGE.
BETTER IN MATH.
DISCUSSES WELL.
TOTAL.

TOTAL TASK DIFFICULTY.

HOME SUPPORT

SUPPORTIVE
PARENTS VERY RESPONSIVE TO NEED
FOR WORK.
PARENTS VERY RESPONSIVE TO NEED
FOR SKILL.
FATHER ESPECIALLY RESPONSIVE.
MOTHER ESPECIALLY RESPONSIVE.
PARENTS ABSOLUTELY ELATED THAT IT
WASN'T ALL E'S.
"D: REALLY TORE THE PARENTS UP."
AUNT AND UNCLE WHO REALLY CARE.
PARENTS WILL BE SURE THE MARKS ARE
A'S.
TOTAL SUPPORTIVE.

PROBLEMATIC (OFTEN LEADING TO POOR
STUDY HABITS)
ALL PARENTS.
DEATH OF A PARENT.
RECENT DIVORCE.
RECENT REMARRIAGE.
LANGUAGE PROBLEMS (SECOND LANGUAGE).
SINGLE PARENT SELDOM HOME.
BOTH PARENTS WORKING, TOO TIRED
FOR DISCIPLINE.
ELDERLY PARENTS WITHOUT MUCH ENERGY.
FATHER LEFT THE HOME, ANGER.
MOTHER HAS HAD SEVERAL HUSBANDS,
NAME CHANGE.
SISTER ON DRUGS, HOSPITALIZED.
TOTAL PROBLEMATIC.

UNSUPPORTIVE
MOTHER RAN HIM DOWN SO BADLY.
MOTHER SAYS HE IS MENTALLY RETARDED,
HE ISN'T.
ABSENCE OR TARDINESS EXCESSIVE
WITHOUT ILLNESS OR EXCUSE.
PUNITIVE, RIDICULOUS PENALTIES.
TOTAL UNSUPPORTIVE.

TOTAL HOME SUPPORT.

CLASSROOM BEHAVIOR/NATIVITY/DEVELOPMENTAL

PHYSICAL
GROWTH SPURT, GROWING RAPIDLY.
VERY LARGE, HEAVY, BIG FOR AGE.
SMALL FOR AGE.
HARD TIME WITH HIMSELF.
PUBERTY.
ON MEDICATION.
CAN'T SIT STILL LONG ENOUGH TO DO
ANYTHING.
TOTAL PHYSICAL.

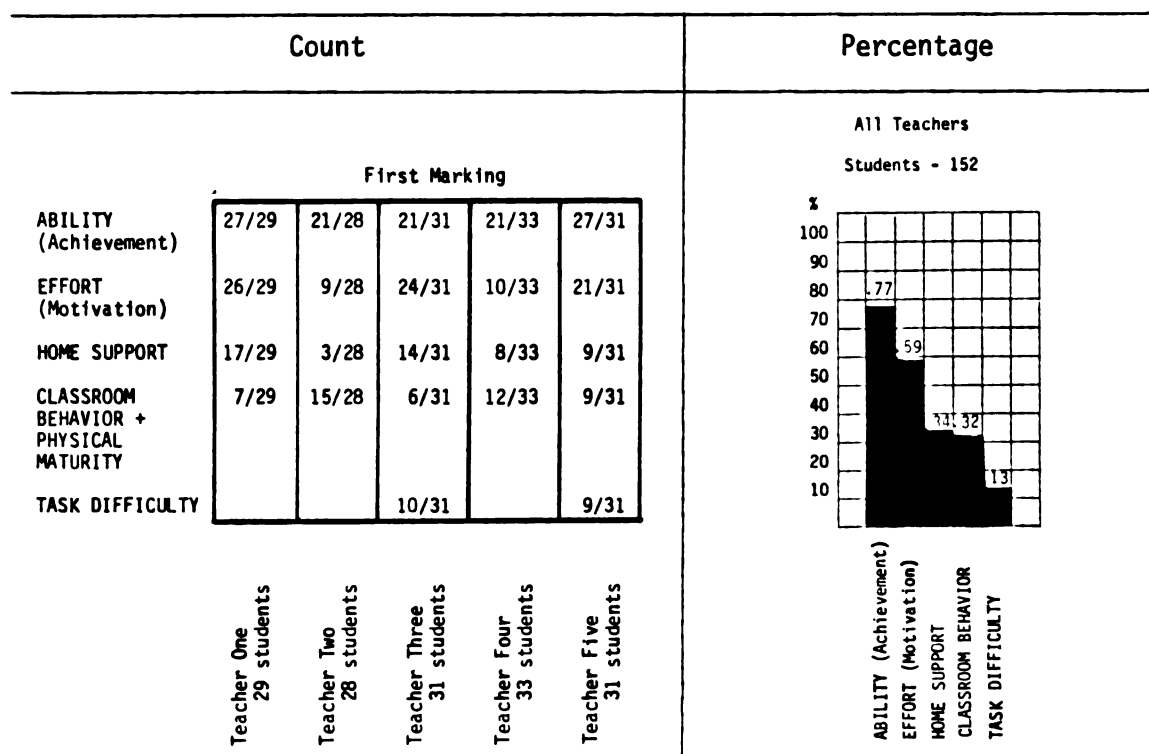
SOCIAL
VERY WITHDRAWN.
MISS SOCIALITE.
INTERESTED IN NAILS, HAIR, ETC.
LADY'S MAN/BOY CRAZY.
TALKATIVE, LIKES TO VISIT.
FLIGHTY, CAN'T SETTLE.
TOTAL SOCIAL.

EMOTIONAL
EMOTIONAL PROBLEMS, PERSONAL PROBLEMS.
VERY, VERY SENSITIVE.
CONSTANTLY WORRIES.
VERY IMMATURE.
VERY MATURE AND DEPENDABLE.
ALWAYS HELPS UNDERDOGS, KIND.
LIKES TO PLEASE OTHERS.
LIKES TO PLEASE ME (TEACHER).
YELLS OUT ANSWERS, LACKS CONTROL.
NERVOUS PROBLEMS.
TOTAL EMOTIONAL.

TOTAL BEHAVIOR.

Table 4.7

Composite Attribution-Utility



Note. The left side of the table displays the actual count of comments made by each teacher within each category against the total class size and of all teachers against the total 152 students. The right side of the table displays the total percentage within each category of all teachers.

All teacher comments pertaining to individual pupil's marks were coded. Once the categories of home support and classroom behavior were identified, no additional miscellaneous category emerged despite the efforts of the researcher to locate unique concerns which did not fit a framework.

Task difficulty did not emerge as a major concern for most teachers and needs further explanation. Teacher problems with task difficulty were frequently solved by their use of individualized programs, grade level reading materials or special education support. Hence typical comments were:

She struggles to maintain Bs. She's sort of a C+, B- kid (coded under effort and ability).

She has learning problems, she's just plain slow, low I.Q. She just has to have everything taught on her level or she doesn't get it. If she were in a traditional classroom where the teacher was teaching all fourth-grade level, she would be in big trouble, but I individualize in basic skill subjects (coded under ability).

She has an A in reading, but I have starred that because she is reading at third-grade level which is two grades below, but she is doing excellent work at the third-grade level (coded for ability and effort).

I do not grade him in math; he is a learning disabled student (coded as ability).

Math may dip as concepts become more difficult. She is a hard worker, a good student, but I tend to feel that she is a low achiever (coded for task difficulty, effort and ability).

She has not mastered multiplication yet, hence, she has a D (coded for task difficulty).

These comments indicate great overlap between the categories of ability and task difficulty due to teacher's choice of words. To clarify categories, task difficulties were coded closely with specific subject matter. This overlap problem is addressed at a later point by combining categories. However, it should not be concluded from this that teachers lack concern with task difficulty.

The categories of home support and classroom behavior also deserve some further explanation. Whereas effort, ability and task difficulty are common attributional categories, the home support level appeared to be an attempt on the part of teachers to qualify and perhaps control effort or lack of it. Not only is the category frequently mentioned by teachers, but it is invoked at both extremes: that of insufficient support or of sufficient support to gain leverage for more effort.

Consider the following comments:

I predict the grade will go up because she has not mastered multiplication and received a D which hopefully woke the parent up and let him know we have a problem.

Mark will remain A. High expectations from home.

She might improve if the mother is interested. It all depends if the mother wants her to write better.

Well, I think mostly because parents start taking a very active role in their children's education when kids bring home bad grades.

In some cases it can damage a child's self-image but on the other hand there are kids who can be motivated by low marks. I find that I not only have to know the child but his parents too, to know if a low grade would be damaging or motivating. And it only takes the first conference to know.

When I got to working with his mother, he finally shaped up.

She is really below average student, but she works extremely hard and has a very supportive father, and I think he is the driving force here.

C-, D+. I predict D and D, split family. Mother works nights, step-father emotional problems. Not a lot of energy left for school work.

C-D. Marks will stay about the same. She lacks self-confidence. Parental support is lacking or was. I hope it changes, but she is late a lot and misses study time.

Home motivation is strong.

Almost no motivation from home.

Comments such as these were coded under home support level and were clearly used by the teacher as a factor in predicting future academic success or failure, an attribution process.

Class behavior was also a category which emerged and needs delineation. This is a utility concept. That is, it was important to the teacher to maintain on-task behavior and to maintain the flow of classroom activities. This maintenance

of flow was a goal in itself, separate from achievement but also related to it (Joyce, 1980). Activities were planned to accomplish academic tasks; completed tasks were equated with achievement. Therefore any disruption of class flow took time away from a task. For individual students who caused distraction, the loss of time was personal, but frequently, if the flow was disrupted, the loss of time was general. Where teachers perceived that sociability, excessive talking and lack of concentration were disruptive to task-oriented behavior, they mentioned these characteristics in relation to predicted marks, i.e., "Her mark will probably go up when she controls her talking." Each teacher stated that they allowed some level of conversation during class, hence, comments on excessive talking, goofing off, teasing, etc., were interpreted as off-task behavior which the teacher was attempting to bring in line. Since the mark was based on the tasks completed, it was assumed that the teacher's comment regarding a low grade was a recognition that zeroes result. Tasks were incomplete, and therefore off-task behavior lowered a mark. The category of classroom behavior lent itself to the decision-tree method of utility analysis.

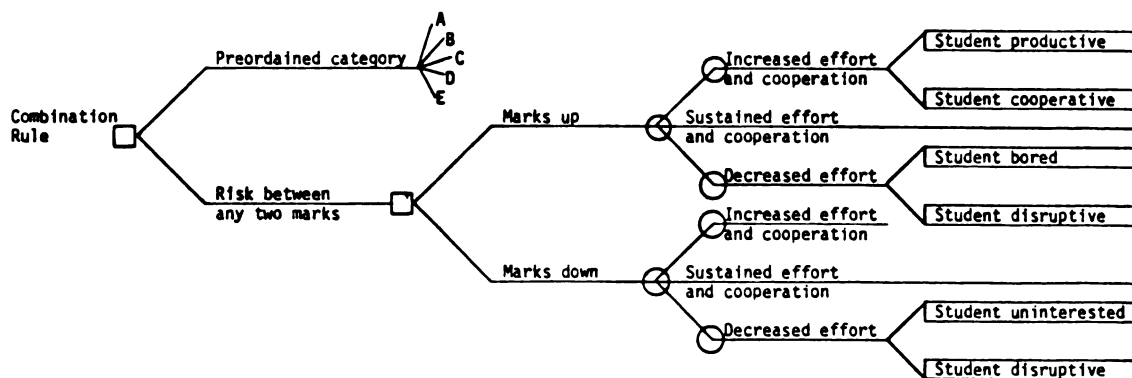


Figure 4.4. Decision tree for marking judgment (adapted from Winstein, Fineberg, et al., p. 18).

The decision tree is the fundamental analytic tool for utility theory. It requires the decision-maker to identify alternative action which might be taken

and to obtain information at these times. An optimal course of action is desired. In the case of teacher marking, the teachers were concerned to have students complete tasks and to maintain class flow (on-task behavior). The decision tree indicates points of risk and the following teacher comments indicate that teachers risked marks to maintain task behavior during class.

Basically a C student but has a behavior problem with another girl in the room, too much talking, but will probably do better now that I have moved her away.

The C will come up I'm sure because she is very bright, but there is too much talking; she is one of my worst talkers.

He's fairly bright, but he's been a discipline problem since he's been here. I am trying something new this week with him. Peer pressure. I assigned a couple of kids to every time _____ goofs, to yell out "_____ get to work." Yell out right in class. It only started today, but _____ knows he's being watched by someone other than me. He just needs that. I look for him to go up, but he's the most distractable kid I have every seen and distracted mostly to bother other people.

She is my second worst talker, but as I get a handle on her behavior, I think the grades will pick up.

Teacher Four gives minuses and pluses in some areas:

. . . if they can stick with the task. If they are half way between an A and B, and they've got minuses in my book because of fooling around in class when they're supposed to be doing experiments or whatever we're doing at the time, then I would give them the lower of the two grades.

Teacher judgment factors, up to this point, have been established through content analysis by coding verbatim responses and counting the totals in each category. Ability emerged as the primary factor and effort secondary. Another means of comparing factor weights, however, is to combine categories which have great overlap and count again. By combining ability and task difficulty and comparing the new category to a combination of the highly overlapping categories

of effort, home support and classroom behavior, a second conclusion appears. Teachers are more concerned with factors related to effort than ability.

Table 4.8
Attribution-Utility Categories Collapsed

<u>Ability/Task Difficulty</u>		<u>Effort/Home/Behavior</u>	
117	Comments Ability	90	Comments Effort
19	Comments Task	51	Comments Home
<hr/>		49	Comments Behavior
136	TOTAL	190	TOTAL

Theoretically, this conclusion makes sense. Teachers, like people in general, perceive that they have more influence over effort than ability (Weiner, 1979). In addition the category of effort covers two purposes of the teacher; individual task completion and classroom on-task behavior. Ability, on the other hand, serves primarily the individual level. The first conclusion which appears from categorizing and coding verbal comments is that ability is the primary judgment cue for marks. The second conclusion, which appears after related categories have been collapsed, is that effort is the primary factor in marking judgments.

The possible third conclusion from the coding data is that the categories of ability and effort are impossible to separate entirely, and therefore continuously present the teacher with a very complex judgment situation. It is difficult to clearly weight one category over another as a judgment cue. It is safe to say that high ability and high effort at the elementary level result in high marks, and low ability and low effort result in low marks. Any other combination of ability and effort presents complexity in isolating judgment cues if one remains within this counting mode of analysis.

There is, however, a different way to analyze the data. When research goes beyond counting comments, one notes that descriptions of unstable effort and distracting or disrupting classroom behavior are much more lengthy and elaborate than descriptions depicting productive students. The longest and most elaborate descriptions involved home support level as a rationale. Contrast the following key descriptions:

Descriptions of brightness with effort.

She is a very bright kid and a hard worker.

Again a bright girl, high expectations from home.

She is an overachiever, a good student and very conscientious.

A, A. Just a very highly motivated student, a lot of internal motivation as well as from home.

B, A. He'll stay there. He just seems really motivated.

He is my brightest, hardest worker, and I predict he'll be an A all year. He does all the extras too.

A, A. Very bright, has had a discipline problem up until now, no more though.

B. A. Very bright but a little disorganized.

B, B. Same. Both of these kids are just kind of steady workers.

Description of uncertainty with unstable effort.

D in math. D in language and I predict they will come up. He has a lot of home influences in his background. He is living with a grandfather, the mother has gone to court to achieve custody. I feel he is a much brighter student than his marks show. So if I can get him motivated myself, I think his marks will come up.

She has a D in math, C in language. This is an underachiever, immature student. I think she is capable of doing more, but she is much more interested in socializing than working. A loving little girl, capable of improving.

Has a D in math, a D in language. I referred _____ for language disability because I believe he is much brighter. He is reading a couple of grade levels below where he should be and yet through social studies discussions, I feel he is quite frightened. He is being tested right now.

C, D. Language will stay, but the math might drop. Again because of the complexity of the math coming up for P. And P probably is pretty much of an average student but very immature, not especially motivated to do well. She is the youngest child in the family which, I think plays some role.

Dropped from A to B in language because of incomplete assignments. I think he'll get it back up. He's a very bright boy. _____ has a lot of personal problems which—he's doing a pretty good job of keeping from interfering with classroom, but he really does have a lot. A single mother with—well they've had a lot of problem with people breaking in the home. He's seeing a psychiatrist too. That seems to be—at the beginning of the year he had a real nervous "tic" in his throat. I don't hear it anymore. He's getting that under control.

C, D. She is my second lowest child and expected to stay the same. Real problem I have with this one, because I've recommended her for testing and her parents refused. Her parents insisted that I put her in a different book because the book I gave her was too easy. So I've had a problem with the parents because they refused to recognize that she is slow. I had two conferences in the same week with the parents. Finally I just let them have their way. I just put a note in her file to the effect that I will not accept the responsibility for the book that she is in. I requested testing, but they refused.

Elaboration of problems is, therefore, a powerful way to weight the concerns of teachers. Using elaboration, descriptions within a category indicated that teachers were much more concerned with ability accompanied by unstable effort than with ability accompanied by steady or extra effort. Effort in these cases was the key judgment factor, not ability. Once again, however, the descriptive data indicated a high interrelationship.

The role of predictions in the marking judgment deserves separate discussion. Does teacher expectancy (prediction) play an important judgment role during evaluation? Many studies reviewed by Brophy and Good (1975) concluded that expectancies became self-fulfilling prophecies in the classic model of

Pygmalion in the Classroom, a well-publicized study by Rosenthal and Jacobsen (1968). The expectancy model holds that "Early in the school year, using the school records and/or observations of students during classroom interaction, all teachers form differential expectations regarding the achievement potential and personal characteristics of the students in their classrooms. Some of these initial expectations are inappropriate, and some are relatively rigid and resistant to change even in the face of contradictory student behavior." These expectancies cause teachers to begin to treat students differently (Brophy & Good, p. 39).

The evidence contained within this study of teacher marking does not contradict this hypotheses, but it questions it. For example, as evidenced in the statistical analysis of the composite study, teacher predictions were higher than actual marks throughout the year. Teachers continued this pattern in the cue sort exercise which was given immediately after the final mark. The cue sort (see teacher cases) was not statistically meaningful, but it emphasized the strong relationship between ability and effort and pointed to continuous high expectations for students. These expectations were born out through the high composite B averages. Looking at the verbal analysis, teachers discussed home support factors and classroom behavior primarily in relation to prediction at the beginning of the marking period. But when the actual summative marking judgment was made, teachers tended to routinely look at work completion factors. The majority of marks were determined by routine procedural rules.

Predictions, therefore, did not appear to determine summative marks. Predictions appeared to be based on the teachers' perceptions of effort to be expended as estimated by knowledge of the attribution categories, whereas marks were based on actual effort expended as measured through task completion in the record book. Optimistic predictions may reflect the teachers' sense of control to influence task completion positively either by leverage through home support, or

by creating interesting and captivating academic tasks. Hence predictions may tell as much about teachers' professional self-confidence as about students' ability and effort. While the evidence in this study does not contradict the expectancy theories, it does suggest that prediction may be quite a separate type of activity from marking judgments and that it might be useful to distinguish carefully between the terms prediction and judgment. Prediction appears to be more closely related to planning and beginning activities than to marking judgments and concluding activities.

From alternative methods of analysis, judgment cues used by teachers in the marking process have been identified and weighted. The inference can be made that although elementary teachers acknowledged ability as a major cause of success or failure, of higher or lower grades, they were more concerned with maintaining individual effort toward task completion and its counterpart whole class effort toward on-task behavior.

One last concern in this study was whether or not teachers used the summative marking process as a feedback mechanism for their teaching effectiveness. The following questions opened the way for responses: (25) What do these marks (formative) represent? (Probe: Why were these specific assignments chosen?) (27) Are there activities which occur which you don't mark? (34) As you look at this whole group of grades, how would you say this group is progressing? Are you satisfied? (3) You have said that you are generally (satisfied-dissatisfied). Will you change any of your plans for next year? (10) Consider the following situation: A year-end marking in which more than three-quarters of the 30 students in Teacher X's class receive C or D. What would you conclude?

Answers to these questions led to the conclusion that summative marks were not a feedback mechanism, however, individually assigned tasks, especially

tests, may serve as such mechanisms. All teachers were generally satisfied with their classes' progress. All teachers ended the year with a class average in language arts and math which hovered close to a B. All teachers were surprised that the average was so high and openly pondered the implications. All teachers rejected the occurrence of the simulated situation in 10 as possible.

I never heard of a situation like that. I can't ever think of a situation where three-fourths would receive Cs and Ds. I would want to know the record of the preceding year, whether they made any progress at all. I'd want to know the reading level. We've never had a situation like that here, and I've taught for long enough that I, usually have at least some really good student and then you've got some that are not outstanding, but they're good. If they're reading below grade level, we write that accordingly. That's why I can't understand, can't imagine that situation.

The teacher is not doing his or her job. Three-quarters, did you say? I can't believe it. You know—that would be my first inclination. I would suspect that something was getting across to that top group of students. I'd want to know what type of unit the teacher happened to be teaching, if the teacher was giving individualized instruction versus group, whether there were any personality conflicts (laughs). I would suspect that right away. Was the teacher having any problems of his or her own, for one reason or another? Especially if I were the principal, I would check into that first. That isn't the way the year should end up. There are too many good students in a classroom. Given the chance, they'll produce for you.

C or D? Well, one conclusion could be that the students were not up to relating well with the teacher and were not really motivated to please the teacher. Another factor could be that there was little home motivation and maybe school is not interesting. I can't really imagine it happening—not here.

In contrast to the rejection of the simulation situation, the teachers responded to question 27 by indicating that task failure was a feedback. They stated that they did not mark all activities. Those most frequently excluded were introducing new material, pretests, class discussions and tests failed by a significant portion of the class. This, therefore, became a direct question in the second interview. Answers included:

For instance, a complete failure across the board? Maybe five people passing the subject. Then I feel that I failed and my grade is bad. So no grade goes in the book. If it has been something disastrous in the social studies area, we just had a study on Canada, and they all did poorly with the exception of three children, we decided to throw that out. And I would reteach it from another approach, and we had a person come in and give us some additional information on Canada, and we are in the process of using a couple of movies. Then we'll go back and try again.

I would reteach if I felt the test was good. If I felt the test tested what I wanted it to test. Then if the kids were not achieving at the level I would expect, as an overall pattern, then I would go back and reteach and then give a similar test again.

I would take it personally to mean that I didn't teach it very well. I didn't motivate them to learn. My response would depend how important I felt the material was. If I felt it was very important, then I would reteach it. If it was something that was, just happened to be in the book, but I didn't consider it important very subjectively, then I would just let it go and sometimes I would give the kids a chance to redo the test or take it home and correct it for some credit.

(Laughs) Either I didn't teach the material or the questions were presented to them in such a way as to be too confusing. So if it happened, I would either strike out the test and not even put the mark in or I would lower my standards for that particular test and change the grading standard.

I have a standard grading system, percentages, 90-100 is an A, 80-90 is a B, etc. The only difference I know is one time during the fall, I changed the grading standard for a social studies test because the highest score was eight out of ten. It was a hard test. In this case, I put eight was an A, and then I just went seven was a B, etc. So I did adjust that one. (This teacher weights tests as equal to homework tasks.)

The contrast between the responses to failure on a test and to the simulated situation where the summative marks were three-fourths Cs and Ds with no outright failures, supported the idea that tasks were more important indicants of learning and effective teaching to teachers than summative marks which were seen as a routine averaging of task marks.

Some anticipated findings did not emerge from this investigation of marking. This may be the result of the interview format or the lack of expertise

on the part of the interviewer to probe at significant points. Specifically the researcher had hoped that teachers would discuss the selection and quality of assigned tasks to a far greater extent and make distinctions between tasks which would indicate different values (weights). Instead, most teachers remained on a general level discussing the significance of the number and variety of tasks rather than the quality. However, most weighted tests more heavily (double) than other tasks. Some assignments were simply checked in or out rather than being corrected, indicating a quality decision. There seemed to be a general assumption that most tasks were worthy, especially the individualized skill tasks which came from published materials and occupied many entries in record books.

Conclusions which could be drawn from this lack of detailed comment are likely to be spurious. However, the discussion is important because it appears to be related to task difficulty which most teachers did not list as an important attributional category during the marking judgment. The conclusion which seems most probable is not that teachers do not make such distinctions, but rather that such distinctions are important during the planning process (beginning) instead of the marking process (concluding). Once tasks are selected, they become part of an assumption which teachers operate from, and which are questioned only at times of failure as when the majority of the class failed a test. This further supports the finding that during the marking judgment, teachers are not concerned with task quality so much as with standard of task completion.

Other anticipated findings did not emerge. Teachers seldom discussed the future placement or success of the student beyond the current year. They seldom discussed their own marks in relation to any outside criteria except the Michigan Educational Achievement Program (MEAP). They seldom discussed standardized testing outcomes as a measure of effective teaching. These would have indicated out-of-classroom measures. The findings which emerged, combined with those

which did not, indicate that teachers concerns during the marking judgment were largely bounded by classroom parameters.

Composite Summary

The purpose of the study of teacher judgment during marking was to create an understanding of the judgment processes which engage teachers across a year and to see if these processes indicate a pattern or model. The data gathered and analyzed have fulfilled the purpose. The findings indicate that teachers generally follow procedural and contingency rules which divide the process into three stages: collection of task completion information, computation of task information and modification strategies to deal with uncertainty between categories and with failure. A marking judgment model, derived from the combined findings of statistical and verbal analysis, illustrates these stages and is included in the integrated summary following the teacher cases.

The findings indicate that the statistical techniques help determine outcomes, hence, they are a signal for caution. The statistical techniques of multiple regression, Pearson correlations and partial correlations, revealed that each marking period made a contribution to the final mark. Regression analysis generally weighted the final prediction and the second mark as the most influential predictors, supporting a recency model of decision making. The Pearson correlations revealed that all marks were more highly related and significant than the regression equation leads one to believe, although the Pearson correlations supported the general weighting patterns of the regression process.

Partial correlations which control for specified variables, questioned the recency effect, and supported an "anchoring" or primacy effect. Both the recency and anchoring effects are well known as common judgment processes. The teachers, however, stated that they strictly averaged the separate summative marks across the year thereby contradicting both hypotheses.

The findings of this study indicate that each marking period stands on its own tasks. Teachers do generally average formative marks at the end of a marking period, and they do generally average the summative marks to arrive at a final mark. However, an analysis of record books, of minuses and pluses and of verbal protocols reveals that they do not do this as "strictly" or as "cut and dried" as they perceive. Instead they have contingency rules which operate in zones of uncertainty and in exceptions. Contingency situations seem to increase as the year goes on.

Teachers have considerable commonality about the judgment cues which operate during contingency zones. The cues include ability, effort, home support level, classroom behavior/physical maturity and task difficulty. Effort constitutes the primary contingency cue with ability close behind.

The composite study reveals that teacher marking processes at the procedural level are related to task completion and at the contingency level are related to factors which promote task completion, especially effort. Interest in the home support level is basically related to gaining leverage to maintain or increase effort. Interest in classroom behavior is also related to maintaining on-task behavior of a significant group of students to assure task completion.

Taken together, these procedural and contingency judgment processes reveal that teachers' marks are task focused and classroom bound.

Case One--Teacher OneBase Data

Sex: Male.

Years of teaching: 14. In this district: 14.

Class Size: 29.

Grade and composition: 6 fifth/23 sixth; 13 boys, 16 girls.

Parents attending last conference: 25 of 29.

Philosophy and Rule

Teacher One believed the current system of marking was satisfactory. During an experience with a Pass, Satisfactory, Unsatisfactory system, he felt that teachers did not have "enough breakdown." He felt that "written comments to parents is the largest controversy in the school because of the time it consumes." Teacher One described his philosophy that "children should be appraised weekly and know where they are on the parameters that I have set: how they compare to their peers, how they are comparing to the material and what they are accomplishing. I talk to them as a class and have them figure out their grade, and if it is a serious situation then I will have a conference." To this end, Teacher One set up operating rules:

Procedural rules:

1. I put everything in the book so I don't have to try and remember. I don't try to do any grades from memory.
2. Roughly thirty grades were considered in the first marking. I try to give roughly a grade per day in each subject if we

cover the subject that day. Some grades reflect three days of work, for instance when we are working on a series of plays and writing a play.

3. The categories in the record book are planned weekly.
4. We start out in the beginning of the year with everyone in the same area, same book, same pages. The ones who excel, the ones who get five 100s in a row then are selected out. They form a fast moving group called the "jet set" but this group fluctuates continually.
5. There are some things I don't mark like introductions to new units and pretests in new areas.
6. I mark in percentages and convert to the marks. 96 to 100 is an A, 90 to 95 an A-, 85 to 89 is B, 80 to 84 is B-, 75 to 79 is a C, and I only give one C and one D category.
7. I weigh a test grade right against daily work. I don't give it extra weight. A lot of people do not test well, and so I would rather down play a test and up play daily work. If the daily work is being done by a parent for instance, then it gives the child an unfair advantage, but it is very very evident as soon as we do take a test. Then I adjust the weighting.

Contingency rules:

1. We also have group grading where six children may receive the same grade for a project. But if one person is fooling around and the rest are working, this person may receive a lower grade.
2. If I have a complete failure across the board, that is, maybe five people passing the test, then I don't record the grade. Then I feel I failed. We just had a study on Canada and they all did poorly, with the exception of three children, so we decided to throw that out, and I would reteach it from another approach.
3. Also, I allow the child to throw out their worst grade. In daily grading I always give them the benefit of the doubt.
4. Some subjects are harder to mark. Math and spelling are easy and right out of the book. Social studies, reading and language arts are difficult to find out precisely where the child is. They involve a lot of correcting of papers and tests. I think they should know something pretty specific, not just generalizations. For instance, if they said there are many provinces in Canada and two territories, then I weigh

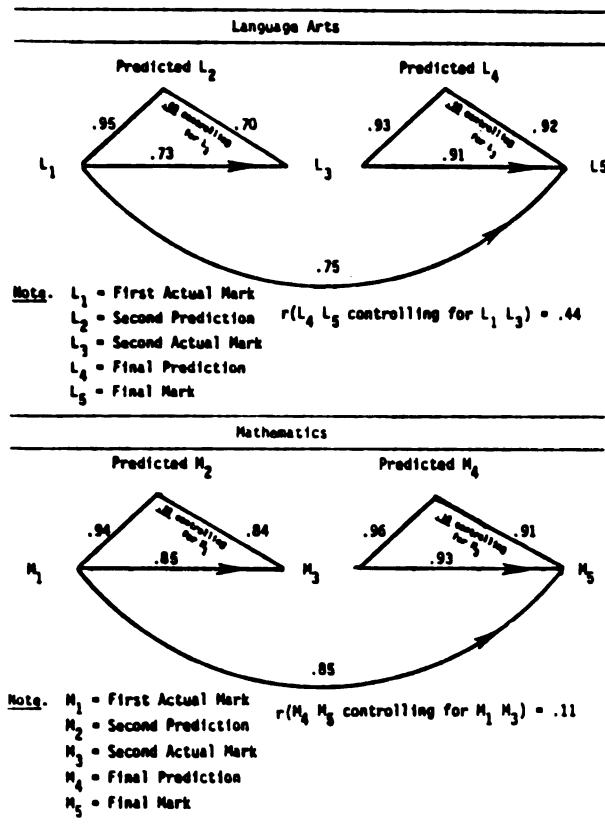
that, okay, what do they mean by many, but if they go on and tell me something about the provinces and they have just forgotten to give me ten, then I have to make a judgment, does this person really know there are ten? The rest of the test may tell me.

5. I retain people. However, I don't keep them back unless they can benefit. These two children are capable of something more, but they need a year of stability to know rules and parameters. I try to build these children up. It is devastating to go on and on and get more and more of things you don't understand. On the other hand, _____ will not stay, because I just don't think this student has the equipment or is capable of much more.
6. Occasionally, I give pluses for extra credit work. If a child completes extra problems or completes a challenge, I give them a plus. When I figure the final grades, if the grade was a 95, I would round it off to 96 because of the extra effort work. So a child that is maybe not the sharpest in the world, but is a very diligent worker receives credit for being a diligent worker. However, the child must score a certain score, and I do strictly take that off the top.
7. Kids must end up on their own merit, and I take strictly the four marking periods, tally those and divide them by four, and wherever that grade falls within the half point, I give them that grade. So the minuses and pluses come in there.

In line with these rules, Teacher One set up a record book system. It is the marks derived from this system that are analyzed statistically in the next section. A section analyzing the interview protocol precedes the case summary.

Statistical Analysis

Three statistical techniques (multiple regression, Pearson correlation and partial correlations) were combined to reveal the marking policies of Teacher One across a school year in language and mathematics. The tables used to modify the original regression equation and yield on adjusted marking model may be found in Appendix D.

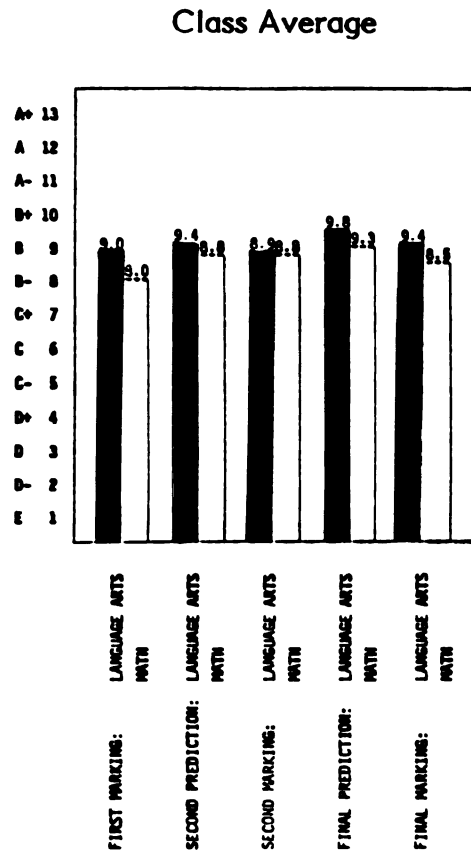


Note. These policies were captured through Pearson correlations adjusted by partial correlations. Summative marks and predicted marks were the base data.

Figure 4.5. Marking policy for Teacher One (with predictions) for 29 students.

The initial regression equation indicated that the final prediction in language (Beta = .60) and the second mark (Beta = .34) were the best predictors of the final language mark. In math, the final mark was best predicted by the second mark (Beta = .69) and no others were significant. The Pearson correlations supported this. Adjustment by the partial correlations reduced the effect of the final prediction of language. The frequency bargraph supported the impact of the actual marks strongly.

Table 4.9
Pattern of Average Marks Across a Year
Teacher One
Students = 29



Note. These averages were derived from the marks and predicted marks of 29 students across a school year.

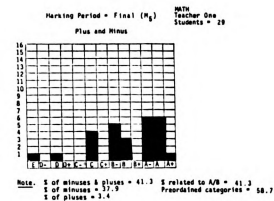
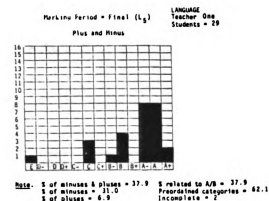
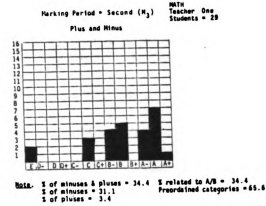
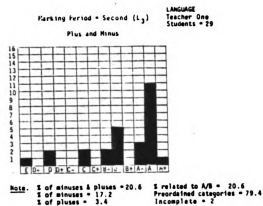
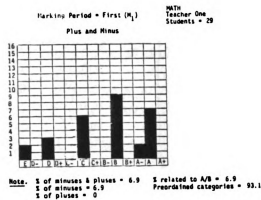
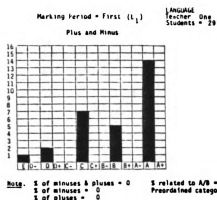
Hence Teacher One's policies resembled the composite model to a remarkable degree. All marks were highly correlated ($P = .001$). Actual marks predicted the final mark more accurately than predicted marks. Predicted marks were more highly correlated with immediate past marks than with future marks.

The frequency distribution of marks for Teacher One also resembled the composite distribution. Teacher One used minuses and pluses, thus depicting the

contingency rules quantitatively. The pattern of minuses and pluses increased as the year continued; increasing from 0 in the first language mark to 20.6% the second and 37.9% the third; increasing from 6.9% in the first math marks to 34.4% in the second to 41.3% in the third. The pattern indicated that all minuses and pluses operate at the A/B level and that there are consistently more minuses than pluses. Despite increasing contingency rules, Teacher One's predominant pattern remained procedurally based with more than 50% of the marks in preordained categories of A B C D E as shown in Table 4.10.

An analysis of Teacher One's record book reveals his task entries (see Figure 4.6). He has a total of 20 tasks which show a wide variety of content including skill dittos, letters, one test, compositions, penmanship assignments and special topics such as Halloween. Percentages are used and averaged into mark equivalencies at the end of the marking period. Teacher One does not use a check-in (✓) system, but he does enter some assignments at 100% if they are completed (note columns full of 100s). Teacher One's marking policies are obviously focused on task completion.

Table 4.10
Distribution of Marks Across Three Marking Periods



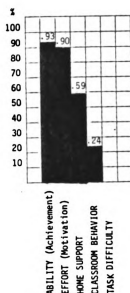
Note. Teacher One's judgment policies followed the general composite model.

Figure 4.6. Record book account Teacher One.

Verbal Analysis

An analysis of Teacher One's protocol indicates that his contingency rules are based on the attributional-utility factors of the composite model.

Table 4.11
Attribution-Utility Percentage
Teacher One
Students 29



The pattern of weighting factors is also representative. Ability (.93) and effort (.90) vie for first influence on judgment. Home support level (.59) is clearly his next concern. His descriptions of home situations are elaborate and closely related to effort which increases the impact of effort as a category.

Anecdotes illustrate the complexity of Teacher One's weighting of cues between categories.

Now I had a parent last year who did all the kid's work; it was very evident. In fact we wrote on the paper, I wasn't the first one, dear Mr. So and So you've done very well on the daily work and sent the work home. I don't know if it ever got there, but we graded the dad as he did the daily work and the kid didn't know anything,

didn't know from applesauce as to what was going on. He had three children in the school and he did all the work; he was very busy. (Teacher graded down, tasks not completed by student.)

_____ is a B- in math and a D in language arts and that is a severe drop. The reason was she just didn't complete a lot of subjects, a lot of things; her mother came in, she tried to make them up, she didn't make of three of them and then just gave up. She's a very moody child. They just told her she was going to move to Arizona and the family's so screwed up you can't believe it. The dad is a 25- year marine sergeant, so you know how that is. There is absolutely, I mean he does ridiculous things like penalizes the kid to the house for 90 days! That's like a court martial you know. That's not using your head in discipline. The kid forgets what they're being disciplined for. Scrub the john for 120 days. Geez, dumb! I can't get it across to him, you know. (Teacher graded down. Lack of task completion overrode home problem.)

I'm going to pass him. It wouldn't do any good to fail him. He's got a dying grandmother living in the home, they just moved her out. His parents are older than I am, older than we are. They're in their late 60s and here's this little squirt coming along. He's got all kinds of physical problems but I think 90% of them are made up. The parents are heavy smokers. He does have chronic chest congestion and allergies to smoke and dogs and cats and all this stuff. He lives in that house you know and then he worries. He's constantly worried his dad's going to die, his mother has attacks, sugar attacks and he comes home and finds her out on a sugar attack, has to go and get her insulin and give her a shot and the grandmother, uh, it tears him up all the time and they do crazy things like call the school and get him ready to come home. No explanation. So then he's blubbering around the room wondering, my mother's dying, or my father's . . . he's always worried, wouldn't do any good to fail him. He could do no better, the poor kids got all of the worries I've got on my back plus a dozen of his own. A little old man already. He had the best, I took him home with me last week and, I take all of the kids home during the year, three at a time. And I took him home last week and this little guy had the best time I think he probably ever had in his life. He was so, he didn't want to leave, he didn't want to come back home and he was so excited he almost wet his pants. He was just shaking he was so excited about going and being there, and I take him down to do a little wood shop project with him and we played games, he was so excited his hands were shaking. Poor little guy. (Teacher graded up. Problems override task completion.)

Despite considerable sympathy with student home problems, Teacher One remains loyal to a task completion criteria. Effort toward that goal must be in evidence or the odds must be overwhelming, such as the quoted case, before Teacher One relinquishes basic procedural rules.

The category of class behavior received very few comments (.24) from Teacher One. Evidently his task orientation does not make behavior a major problem. However, there is not sufficient data to draw conclusions.

Teacher One did not comment on the simulation question because the interview ran short of time. However, he commented emphatically that a failure of the majority of students on a unit test would provide an occasion to reteach. He gave an example of a recent occurrence which was quoted in the composite case. He saw such task failure as a feedback mechanism.

Teacher One resembled the composite case in his optimism. His predictions were higher than actual marks across the year. His class average was high. In addition, the cue sort technique given at the end of year reinforced this conclusion.

Table 4.12
Cross Tabulations of Effort and Ability
Teacher One

		EFFORT		
		High	Low	
ABILITY	High	21	4	25
	Low	0	4	4
		21	8	29

Criteria from outside the classroom did not enter Teacher One's regular marking judgment. "The only time I ever do this (compare students to standardized tests) is if I am challenged on the grade I am giving. I never look at their past record when I give them. The only record I look at is a reading score card that comes with them . . . we try to put them into the proper book."

However, Teacher One does have curiosity about his students in seventh grade. "They send back the testing machine runoff and I compare it just to find out where my kids went, because I like to compare them against past years. They've always done well. Yes, I am finding out how well I'm doing." These comments indicate a definite interest in the student's future, but there is not specific detail involved perhaps because he perceives the situation as successful. These general sentiments, however, were never expressed in relation to any of Teacher One's marks or predicted marks.

Summary

Teacher One's judgment policies closely resembled the composite case. He marked the majority of students according to procedural rules and a minority according to contingency rules. His judgment cues related to contingency rules also mirror the composite with more weight for effort and ability.

Teacher One's policies have coherency. They are focused on task completion on a variety of tasks besides skills. His marking is classroom bound.

Case Two--Teacher TwoBase Data

Sex: Male.

Years of teaching: 19. In this district: 17.

Class size: 28.

Grade and composition: 4/5 split.

Parents attending last conference: 26 or 28.

Philosophy and Rules

Teacher Two was very satisfied with the current grading system (see report card in Appendix B). Only as a child had he experienced a different system of excellent, satisfactory and unsatisfactory. He appreciated the fact that the school district did not have a specific philosophy of grading and felt "thank goodness, I hate to be dictated to. I like it just the way it is because we all have our own criteria for grading."

Teacher Two "see grades as having two purposes: one to inform parents one to motivate children. Not to inform children, because I think children in my class always know where they're at anyway. But sometimes I grade down in order to motivate them to try harder next time. But my primary thing is to report to parents."

To this end, Teacher Two has certain rules which he applies to marking. Some of these rules or strategies are embedded in philosophical statements.

Procedural rules:

1. Our report cards give us a chance to put down the grade

level at which students operate as well as grade. Since I teach all basic subjects on an individualized basis, I have kids reading as low as beginning third and as high as sixth. I could give a third grade reader an A if they were doing a lot of work and trying very hard. In other words, an effort grade. This also works in the reverse. I do have some children that are in the fourth grade in a sixth grade book and getting a C or a D on their report card because they did not read a lot and did not do their workbook properly.

2. I grade according to what we've done. We may have as few as four grades in a marking or as many as ten. If I have less than four, I probably wouldn't even give a grade. I would think the grade wouldn't even be valid. I don't give many grades so they tend to be very important when I do give them, especially tests. Well, I think it's very unfortunate when you only have one test in a marking period to grade, which was the case this last marking period and that was my fault. I just let things go too long and it came to the end of the year, end of the marking period and I didn't have enough time. I think to be fair with kids and give them a chance you should give at least two or three other minor grades to average with those tests.
3. Generally I just follow the old standard: 90s is an A, 80s is a B, 70s is a C and 60s is a D, below 60 would be an E. I weight tests and written reports heavier and usually put them in a different color so they can be seen.
4. In reading, math or penmanship, they can sort of go at their own speed. For instance they know they are expected to do at least one penmanship paper a week, ten papers per ten weeks. Those that had 10 pass papers got a middle B. If they did considerably more they could get an A. Two children got A+s because they had done 35 or 40 papers when 10 were expected.
5. My records are free, they are on my desk all the time. Kids are always looking over my shoulder comparing even competing with each other to see who has the most math papers, etc.

Contingency rules:

1. Well, two things influence whether I'll go up or down. The child itself; if I feel they're trying their hardest, I tend to give them the higher grade. If I think that they are capable of even better than that highest grade, a higher grade above that, I would tend to give them the lower grade to sort of encourage them to try harder next time. And the other thing; earlier in the year I tend to give the lower grade,

later in the year I tend to give the higher grade because I think grades can be used for motivation and you motivate a child more, I believe, by giving them the lower grade to begin with and give them something they can shoot for. That is, if we're talking C+s and B+s but if we're talking very low grades like the difference between an E and a D-, I would tend to go the higher grade if it's in the low range, say below C because it tends to help your self-image and some kids, once I have seen kids just blossom and bloom when I've given them the higher grade, the C instead of the D. And it improves their self-image and they feel so much better about themselves. But when we're talking above a C then I tend to give them the lower grade early in the year and the higher grade later in the year.

2. I don't usually grade daily discussions. If we read out of a book and have a discussion about something, and if the class is cooperative, I don't usually grade. But if they get so they are not listening or paying attention which has not happened this year yet, then I start giving oral reports which I record as pluses or minuses.
3. Marks are a motivating factor mostly because parents start taking a very active role in their children's education when the kids bring home bad grades. This is greatest during the first two markings, but along about the third marking period it's spring and everyone's attention is sort of turned away from school, including parents. Well, in some cases it can damage a child's self-image but on the other hand, there are kids who can be motivated by low marks. I find that I not only have to know the child but his parent, too. To know if a low grade would be damaging or motivating to them. And it only takes after the first conference and even asking other teachers about certain kids to find out whether low grading would damage them or not.
4. Well, like all teachers I have my expectations of what a child should do in my room and I guess I feel that it's my responsibility to not only just give out work and put grades in the grade book, but see to it that a kid achieves up to my expectations. Push, push, push, I'm always pushing them. I'm always reminding them. I'm always trying if they fall below on a test as a class tending to give them an opportunity to do extra credit work to make up their grade. I don't like to send Cs and Ds and Es home. But I do. You looked at my report cards. You see that I do send Cs and Ds and Es. But I give them all the opportunity I can. I'm a mother hen. I really keep after them and try to give them every opportunity.
5. If the majority of the class failed a test, I would take that personally to mean that I didn't teach it very well. I didn't motivate them to learn. Sometimes I would give the kids a

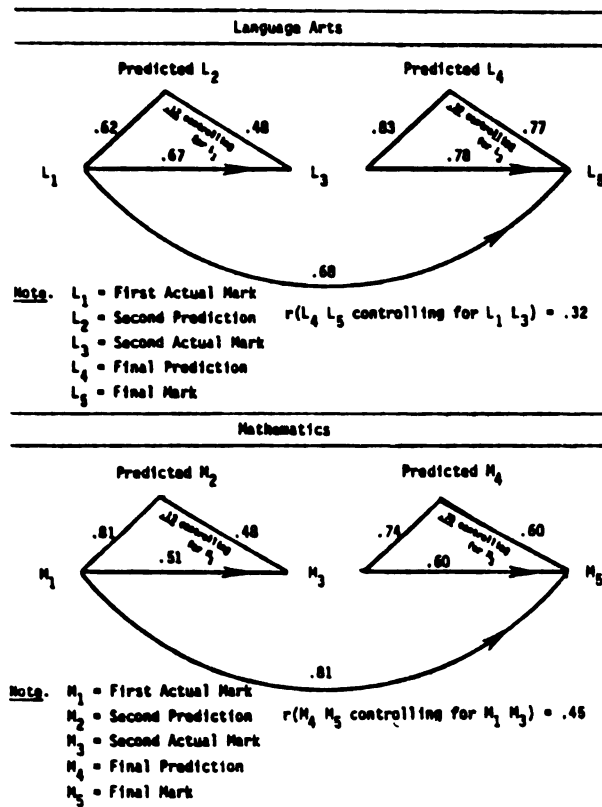
chance to redo the test, take it home and correct it or correct it in class and try to improve their grade or as I did this last test in social studies I had nine kids get an E on the big unit test and I told them I was sorry that the information was taken directly from the book and they were told what the test would be on, they were told that most of the test would be the vocabulary words that were in dark print and all they had to do was . . . so I felt that the test was fair but so many did get poor grades on it and it was right before card marking and I had to grade their report cards the nine kids would have taken home Es and I didn't want to send Es home on report cards to parents. I ended up sending about three home. I gave the kids an option of doing some extra credit questions in their book, strictly on their own, and I told them where the questions were and how to do them and quite a few came in and I was able to raise their grade. It was sort of a band-aid approach to the situation. I sure don't like to send home Es.

In line with these rules, Teacher Two set up a record book system. It is the marks derived from this system that are analyzed statistically in the next section. A section analyzing the interview protocols precedes the summary.

Statistical Analysis

Three statistical techniques (multiple regression, Pearson correlations and partial correlations) were combined to reveal the marking policies of Teacher Two across a school year in language and mathematics (see Appendix D for background tables).

Marking Policy (with Predictions) Teacher Two

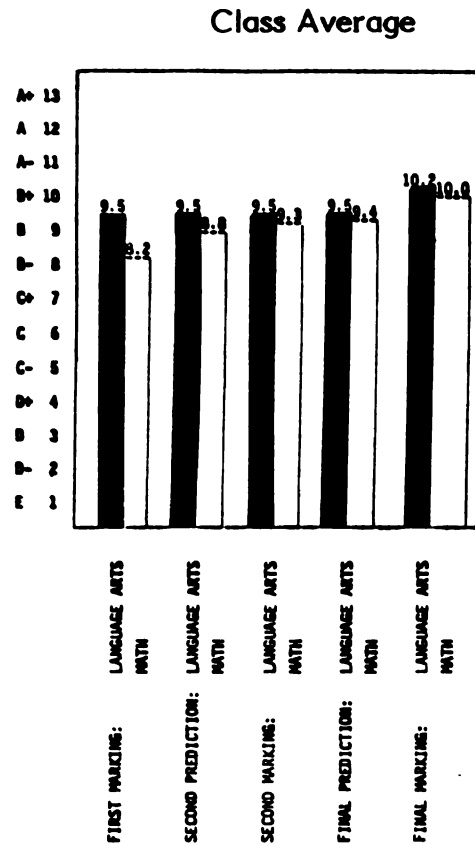


Note. These policies were captured through Pearson correlations adjusted by partial correlations. Summative marks and predicted marks were the base data.

Figure 4.7. Marking policy for Teacher Two (with predictions) for 28 students.

The initial regression equation (see Appendix D) indicated that the second language mark (Beta = .34) and the final prediction in math (Beta = .40) were the best predictors of the final mark. This was supported by the Pearson correlations but adjusted somewhat by the partials. The frequency bargraph appears to be the deciding factor in this policy.

Table 4.13
Pattern of Average Marks Across a Year
Teacher Two
Students = 28



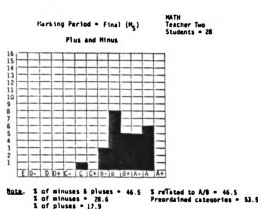
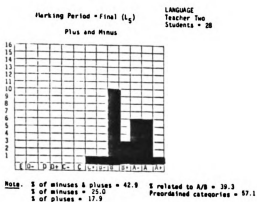
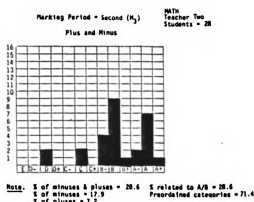
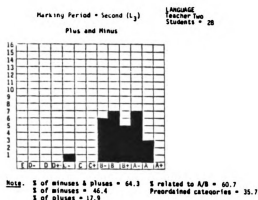
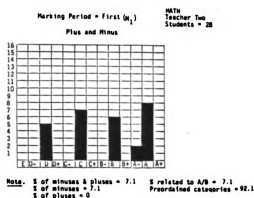
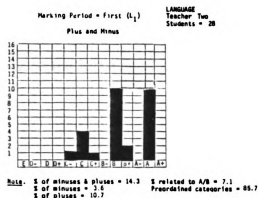
Note. These averages were derived from the marks and predicted marks of 28 students across a school year.

Teacher Two has a 9.5 class average and predicted average in language across the year until the last mark when it increases to 10.2, the highest of all five teachers. In math, the average steadily increases across the year for both actual and predicted marks; from 8.2 to 8.8 to 9.3 to 9.4 to 10.0. Yet these marks result in the weakest correlations of any of the five subjects. This appears to be related to the fluctuating distribution of minus and plus marks. Nevertheless, the correlations are generally high and significant.

Hence Teacher Two's policies resemble the composite model in that the average remains stable despite fluctuations in minuses and pluses. Predictions are more highly correlated to the immediate past mark than those of other teachers.

The frequency distribution of marks for Teacher Two resembles the composite in having minus and plus categories. These categories also follow the composite pattern of having few in the first marking (7.1% in math, 14.3% in language) increasing in the final mark (46.5% in math, 42.9% in language). During the second language marking however, Teacher Two gave 64.3% of his marks with plus or minus modifications which did not follow the composite pattern. Obviously, Teacher Two used contingency rules regularly as depicted in Table 4.14.

Table 4.14
Distribution of Marks Across Three Marking Periods



An analysis of Teacher Two's record book indicates the task entries (see Figure 4.8). Teacher Two had the least entries and relied on creative assignments (writing and stories) rather than skill areas which may account for the use of contingency rules. Teacher Two also puts the greatest stress on literature and book reports. Students gain contingency points through the black dots which appear on the left side of the page.

1st period
book reports

W.
P. 100
F.

194 Period

2nd Period

[illegible]

Note. This record book page was extracted from data collected on Teacher Two's marks in language and mathematics. This depicts marks in language for two marking periods.

Figure 4.8. Record book account Teacher Two.

The conclusion appears that Teacher Two differs from the composite model by specifying fewer procedural rules, by entering less tasks and by allowing considerably more leeway for individual student initiative. Teacher Two's class average resembles the model in being considerably above average.

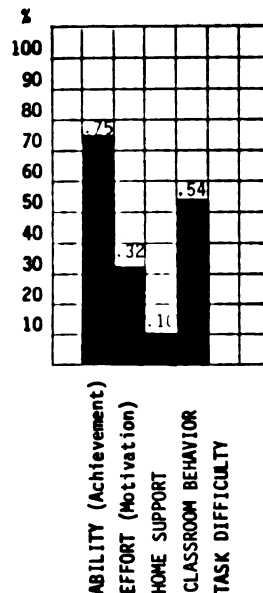
Verbal Analysis

Teacher Two's attribution-utility chart reflected the same general categories of concern as the composite judging model. However, he put different weights on the cues: ability (.75) running significantly higher than effort (.32); classroom behavior (.54) higher than any other teacher in the study; home support comments running lower (.10). Teacher Two seldom commented on task difficulty perhaps because he used an individualized approach in almost every area.

It is difficult to characterize Teacher Two's task orientation. He has fewer tasks in the record book, but he offers the greatest opportunity for students to do even more tasks on a contingency basis. Some students take advantage of this, hence, his generally high class average. The procedural base is less clear. The highly individualized approach may explain the lack of discussion about a specific level of achievement for the class.

Table 4.15

Attribution-Utility Percentage
Teacher Two
Students 28



Teacher Two has some comments and anecdotes which highlight this situation.

I teach all the basic subjects on a individualized basis so I have kids as low as beginning third and as high as sixth.

Usually my problem is not the letter grades I give out so much as the grade level in those subjects especially reading and maybe spelling.

I have usually two to three parents every marking period or every conference period who will wonder how I can give their child an A in reading when they are a year behind in reading. So I have to explain that it is an effort grade. Once they understand that my grades and most subjects are effort grades, they are satisfied.

____ C. D. One of my lowest students and expected to stay the same. Real problem I have with this one because I've recommended her for testing and her parents refused. Her parents insisted that I put her in a different book because the book I gave her was too easy so I've had a problem with parents there because they refuse to recognize that she is slow. Finally I just let them have their way. I just put a note in the files to the effect that I

would not accept the responsibility for the book she was in . . . I suggested testing but they refused.

In fact I had some parents once that refused to have their boy in my room. I said, _____ (principal) I want that boy in my room because they got their older girls through and they were never in my room. I always looked forward to having them, they were nice kids. I finally decided that I wanted this boy, I wanted one of these kids in my room and they were just adamant because they had heard rumors that he doesn't know what he is doing and so on and so on. So I got them in here for conference, this was like the spring because of coming fall and I showed them my records and showed them my way of doing things and they were very surprised and they said okay, well _____ can be in your room. Well, _____ was in my room the following fall in a fourth grade 4/5 split as a fourth grader and the next year I taught 4/5 again and they requested him to be in my room the second year. They were afraid because my system was so different from anything they had before and they thought they had heard rumors that I let kids do as they want to. One reason it happened was because—my kids are working in different pages in different books and they help each other out—there is a higher noise level in my room than some other rooms especially when we are doing math. Math is kind of a noisy time and it lasts about forty-five minutes a day. At that time we had a lot of parents working in the library so they would walk by my room and see all this commotion and moving around, and occasionally I'm sure they would see kids goofing off too, because boys especially will throw a paper wad or misbehave, and of course, eyes tend to focus right there with that one kid doing something wrong, parents don't realize that when you keep kids in their seats and quiet supposedly, they are still goofing off but they are just doing it quietly. They are playing with their little cars behind their book and they are reading a magazine or a comic behind their book, but when you turn them loose to be free, the behavior becomes more overt and it's easier to spot and they spot it.

These extended comments by Teacher Two indicate a problem area in communication which stems from his procedural base. A conclusion present in this case is that Teacher Two differs from the composite model in the extent to which he functions on the contingency rule basis. Most teachers in the study, especially those emphasizing skills, operate predominately on the procedural level for the majority of their marks which is much easier to communicate. Teacher Two is likely to continue having some difficulties in explaining his system to parents.

Teacher Two resembles the model in his general willingness to discard marks if the whole class failed, seeing this as his own failure.

In the cue sort, Teacher Two expressed an average optimism.

Table 4.16
Cross Tabulations of Effort and Ability
Teacher Two

		EFFORT		
		High	Low	
ABILITY	High	13	2	15
	Low	4	9	13
		17	11	28

Summary

The judgment process of Teacher Two differs substantially from the model. In a significant number of marks, he follows contingency rules rather than typical procedural rules. In effect, his students have more opportunity to move ahead, setting their own standard. On the other hand, he does not give as many As as some other subjects.

Teacher Two's judgment policy is very much influenced by attributional-utility factors although he weights them differently, a trait common across subjects except for effort.

Lastly, Teacher Two defines his tasks differently and chooses fewer of them for the record book. In this manner, it is difficult to conclude that his marking judgment is geared toward task completions at a specified level or standard. Completion in and of itself appears to be a more dominant goal. Despite this question, Teacher Two's marking is classroom bound.

Case Three - Teacher ThreeBase Data

Sex: Female

Years of teaching: 16. In this district: 15.

Class size: 32 (5 L.D.) One student moved.

Grade and composition: 5th grade; half boys/half girls.

Parents attending last conference: 29 of 32.

Philosophy and Rules

Teacher Three was basically satisfied with the marking system (see report card in Appendix B). She felt "the conference with the parents is the most important part of the program so to speak, because I do explain to the parent how I mark and why. It's a way for the student to know how he is doing in regards to expectation."

"I think marks are definitely important to the parents in this area, because we hear if somebody's mark is not where the parent thinks its going to be. They usually contact us right away. Most of the parents in this area are successful people and they want their children to be. In turn, they feel that success in school leads to a successful life, and therefore that's their way of measuring how their child is doing."

Teacher Three believed that marks were a periodic summary of student learning and that parent knowledge of the situation brought motivational support. Hence she had rules to support this:

Procedural rules:

1. The record book is very valuable. The marking allows me to summarize my thoughts about the child and how he is progressing, so when I go into the conference I know exactly where the child is in regards to my expectation for him. I have been caught in the spot where parents come in and say, "quick I want a conference about my kid," and I really haven't had time to look back over marks, and get my thoughts together.
2. I keep track of daily work, which I don't always record. I do check a paper daily and I do look at progress daily, but I don't always record it. It depends on the work itself. I may have as many as 20 marks in the book or as few as 10.
3. Marks in the record book represent homework, skill study sheets, workbook pages, projects, oral projects, tests—a great variety. In math, I have a separate page where I list the skills and as students master the skill at 80% or better, I give them a mark. Since they also work in groups, they may change groups.
4. If the majority of the class failed a test, "I would reteach if I felt the test was good; if it tested what I wanted it to test."
5. The first marking and conference are especially important. Until then, the student really doesn't know how you're going to mark. Once he sees what your standards are, then he tends to work harder or less hard, depending on how he felt he worked in comparison to the mark.
6. I don't really look at permanent records until after I have marked the first marking period. I like to get my own feelings for the kid. I do look at the Michigan Assessment tests and they are useful. I pretty much use them to basically support my own opinion. I still size up my kids myself first.
7. Most categories of the record book are planned in advance. However, if we don't complete the task successfully, I may not record it until I have retaught it the next day.
8. I use one very useful grading device which helps keep my marks less biased. It is a chart which figures points into letter grades according to the standards 90-100 = A, etc. That way, I can have six items or thirteen. They don't have to be an even ten or twenty.

Contingency rules:

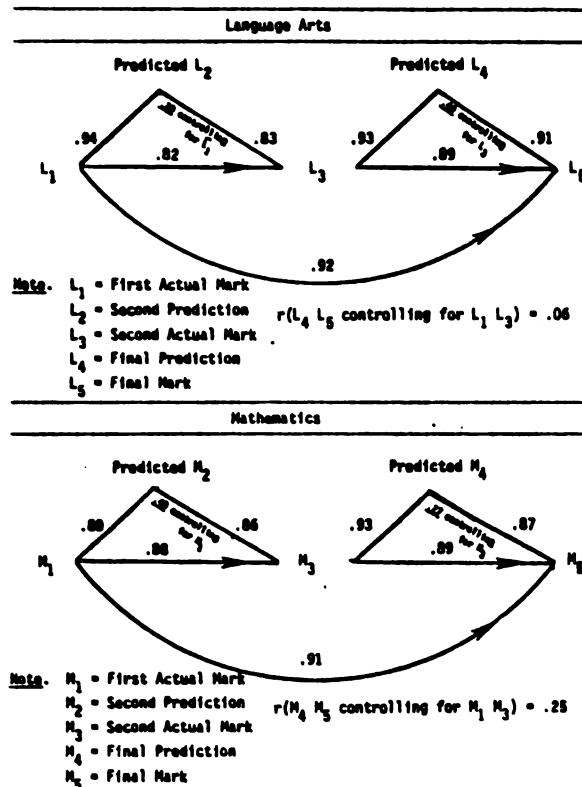
1. If I am in doubt, I look back at the marks to see the nature of the assignment. If, for example, in reading the comprehension marks are all good, but they drop in skill papers, I would probably go with the higher mark. If comprehension marks are low, then I might stay with the lower mark In language I would look back, if the creative writing assignments are the high areas and the low areas are in the skills, I'd probably keep the mark at the low level. I think creativity is important, but I think the language skills have got to be there.
2. I give four points for A, three for B, two for C and one for D, and I show incomplete with circles. I add them up and divide by the total number of marks. Very cut and dried. If the mark is in the middle, I look at the record book and the kid. Is he really working or not putting forth much effort?
3. I sometimes give Ds to make a point with the student and parent. If I slide them through on a C, nobody gets too upset, but a D really bring the point home.
4. Tests are important. I guess I really feel, especially the ones I make up, do reflect what I'm teaching and whether or not the student is mastering it. (I double-check my results with those of the district's criterion reference measures as well.)

I don't give semester tests as such. I test after I've taught so that they come throughout the marking period. It's the one way I can make sure that they are mastering the skill, not just parroting back a short-term learned kind of thing. And I give review tests in skill areas such as multiplication. And again, if the grade is coming out in the middle, if test scores are high, then I'll move the grade up. If low, I'll drop it back.
5. Skill tests are important. I feel fifth graders should have mastered multiplication. If they hadn't, they got a D, which hopefully woke the parents up and let them know the problem.

In line with these rules, Teacher Three set up a record book system. It is the marks derived from this system that are analyzed statistically in the next section. A section analyzing the interview protocol precedes the summary.

Statistical Analysis

Three statistical techniques (multiple regression, Pearson correlations and partial correlations) were combined to reveal the marking policies of Teacher Three across a school year in language and mathematics. The tables used to modify the original regression equation and yield an adjusted marking model may be found in Appendix D.



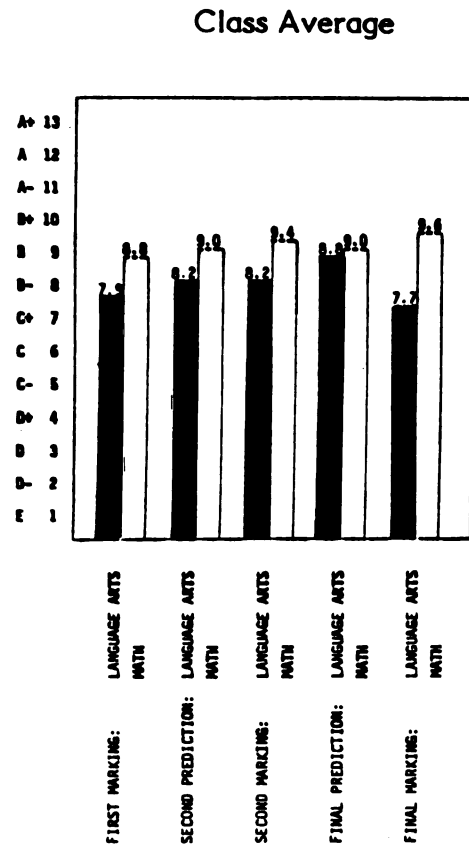
Note. These policies were captured through Pearson correlations adjust by partial correlations. Summative marks and predicted marks were the base data.

Figure 4.9. Marking policy for Teacher Three (with predictions) for 32 students.

The initial regression equation indicated that the first and second mark were the best predictors of the final language mark. The first and final predicted marks carried the greatest weight in the final mathematics mark. Adjustment by Pearson and partial correlations resulted in displacement of the final predicted

mark in math by the second mark (see Appendix D for additional tables). This adjustment was supported by the frequency bargraph depicting Teacher Three's class average pattern. This showed that predictions did not correlate with marks as highly as marks correlated with each other.

Table 4.17
Pattern of Average Marks Across a Year
Teacher Three
Students 31



Note. These averages were derived from the marks and predicted marks of 31 students across a school year.

Teacher Three's policies resembled the composite model. The conclusion appeared that her marks were more correlated to actual marks than to predictions. Predictions were also more closely correlated to immediate past

marks than to future marks. These conclusions supported the teacher's statement that she generally averaged the summative marks across the year, although not quite in the "cut and dried" manner she perceived. Teacher Three has the highest correlation of the individual cases between the first and last marks (language .92 and math .91) supporting a strong relative position on "cut and dried" average.

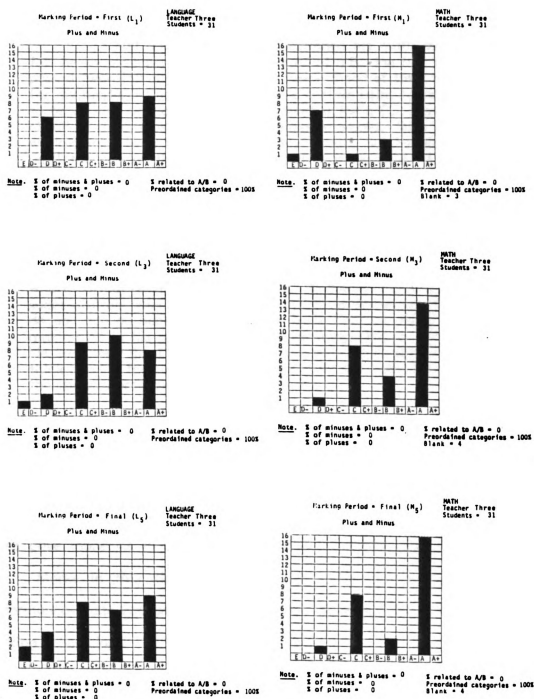
The frequency distribution pattern of marks across the year for Teacher Three indicated very little change of class average or mark distribution across the year and no use of minuses and pluses. This finding further supports the strength of the first marking as a stable unit in the individual regression equation which differs from the composite regression pattern. There is an interesting note, however. Teacher Three gave a disproportionate number of Ds the first marking to make a point that the multiplication tables had not been learned. Presumably this lowered the class average to make a point to the parents. When followed up, this turned out to be the case. The math class average increased steadily through the year, 8.8 to 9.4 to 9.6 as shown in Table 4.18.

An analysis of Teacher Three's record book showed the task entries (see sample Figure 4.10). Of 17 language arts entries in one marking 15 were corrected and two checked in (✓).

Assignment		M T W T F S S							M T W T F S S							M T W T F S S							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story																							
Story				</																			

Table 4.18

Distribution of Marks Across Three Marking Periods



Verbal Analysis

Teacher Three discussed skills and mastery more than any teacher. It is therefore interesting to note her attributional-utility comments compared to the composite case. The category of ability comments was lower (.68) than effort (.77) but task difficulty comments were higher (.32) than those of most teachers. This verified her focus on skills as a criteria for marking.

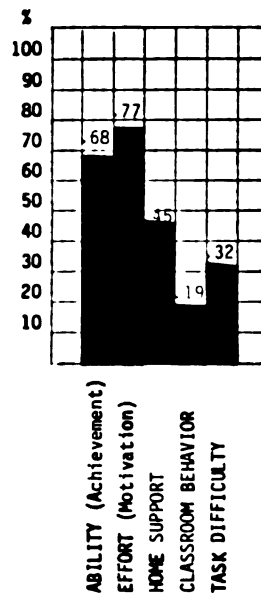
I had one student who questioned his math mark and he felt he should have gotten an A instead of a B. He felt that he had passed two-digit division. The kids knew ahead of time the criteria for my marks in that they had to be a certain point to get an A, had to be at another point to get a B, another for a C. They knew this ahead of time, must have been a couple of weeks. So a lot of them worked very hard to pass certain test levels, but he didn't pass it until the day of the conference which was too late. I was really sorry, but the mark was in.

I will stick with small group instruction. We just finished a geometry unit and I work with the whole class in that type thing. Then I go back into small group or specific skill development. I work with each group daily. I feel that immediate classroom contact is important to me, then I know where I need to reteach the next day, where I'm getting across. Usually it makes the kid feel very successful because I worked with him and he knows exactly what to do and can go back and do it. He feels like he's doing it, always making As in his math even though he is working at a lower level. He feels very good about himself and I think that's the whole key. I think if a kid feels good about himself and I can keep him motivated, then he will progress.

Teacher Three seldom discussed class behavior in the protocols, and the attribution-utility charts attest to this (.19). Evidently, the task orientation which she clearly spelled out to students kept this group of students occupied. Teacher Three gave more As than any other teacher indicating that students were on task. Marks in this class were, therefore, directly tied to tasks and the performance-grade exchange was carried out. Sixteen As were given on the first marking in her class.

Table 4.19

Attribution-Utility Percentage
Teacher Three
Students 31



Teacher Three cooperated with Teacher Four. They regularly compared notes, and neither used minuses or pluses on report cards. Both stressed tasks, both used a marking device (wheel) which allowed them to clearly turn number points corrected on a paper into A B C D E. This allowed them to use uneven numbers quickly.

Teacher Three commented frequently about effort, and rated it above ability, but the idea of effort was directly translated into skill mastery at a stated level of difficulty. In the manner of behavioral objectives as discussed by Wrinkle in his report card study, Teacher Three's approach resulted in clear categories of marks within marking periods. Clear categories, without plus or minus modifications, resulted in generally high correlation across the year.

Teacher Three preferred not to use outside standards to assess students.

In comparing my marks with the student assessment scores or any other standardized tests, I don't really, I try not to let those influence me. In fact, I don't really look at permanent records until after I have marked the first marking period. I like to get my own feeling for the kid I do look at MEAP scores and they are useful.

I pretty much use these tests to basically support my own opinion. I still size up my kids myself first.

No, I don't really look at these tests first. I don't pull out all the test scores and look through the record and say well this kid has always been a B student so I assume that is where he is going to fall, or this kids been the pits all along. I really like to get my own rapport established with the kids, I don't like to categorize them right away. In fact, we get reading scores and math scores for reading groups and math groups from the previous teachers, and I don't even rely on them.

These brief responses are not sufficient for drawing specific inference, but they add to the conclusion that Teacher Three's criteria of judgment about kids' ability and effort was contained primarily within the classroom. When combined with the total lack of comment on any item related to future student work beyond the present fifth grade, it supported the general composite conclusion that the marking judgment was classroom bound.

Teacher Three put heavy emphasis on the home support level in protocol statements and in the attribution-utility chart. She used home support as a leverage to maintain or increase effort on task completion. She did not use home knowledge to modify marks directly with a minus or plus. Teacher Three resembled the composite case in her optimism about her students. Not only were her predictions high, but the cue sort given at the end of the year was also particularly high in light of having five special education students.

Table 4.20
Cross Tabulations of Effort and Ability
Teacher Three

		EFFORT		
		High	Low	
ABILITY	High	15	5	20
	Low	3	7	10
		18	12	30

Summary

Teacher Three's marking policies and judgment cues resembled the composite model. Statistical analysis supported high correlations between marks with predictions having lower correlations. Marking distributions revealed clean categories with no minuses or pluses which differed from the composite. In verbal analysis, Teacher Three considered the composite judgment cues of ability, effort, home support, class behavior and task difficulty. However, she weighted task difficulty more heavily than most other categories which supported her emphasis on basic skill objectives. Teacher Three focused her marking on task completion at a given level of difficulty, and her marking policies were classroom bound.

Case Four - Teacher FourBase Data

Sex: Female.

Years of teaching: 14. In this district 14.

Class size: 33

Class grade and composition: Fifth grade; 16 boys and 17 girls.

Parents at last conference: 29 of 33.

Philosophy and Rules

Teacher Four was satisfied with the current marking system (see report card in Appendix B). Having used others, she stated that "the one we are using now is the best one we have ever had." Teacher Four relied on the conference to make an important two-way communication where she also gets input. She felt that "traditional A B C marks within the conference setting are the best. It's black and white. Probably this is because the parents grew up with it. However, it is very important to indicate the grade level at which the student is being marked."

In light of her philosophy, Teacher Four established procedural and contingency rules.

Procedural rules:

1. My record book contains marks for homework, special reports and tests. I don't count class participation or discussions because I expect them all to participate.
2. I use a point system where A is 4, B is 3, C is 2. I use a calculator across 10 to 20 items and find a strict average for the marking period. In grading papers, I use a percentage system where 90-100 is an A, 80-90 is a B, etc. The only difference I know was a time I changed the scale because the highest scores were so low. I did adjust

that one, and I would be tempted where the test is very difficult.

3. Homework and tests come out the same. I mean I don't weight one over the other.
4. In math, marks are based on a mastery level. If they were into 2-digit division, they got a B. If they are beyond 2-digit division, they got an A.
5. Achievement should have more weight than effort in a grade, but I try to help the child out in other ways. When it comes to grades, the only thing I take into consideration is total difficulty of the subject such as our social studies for fifth grade.

Contingency rules:

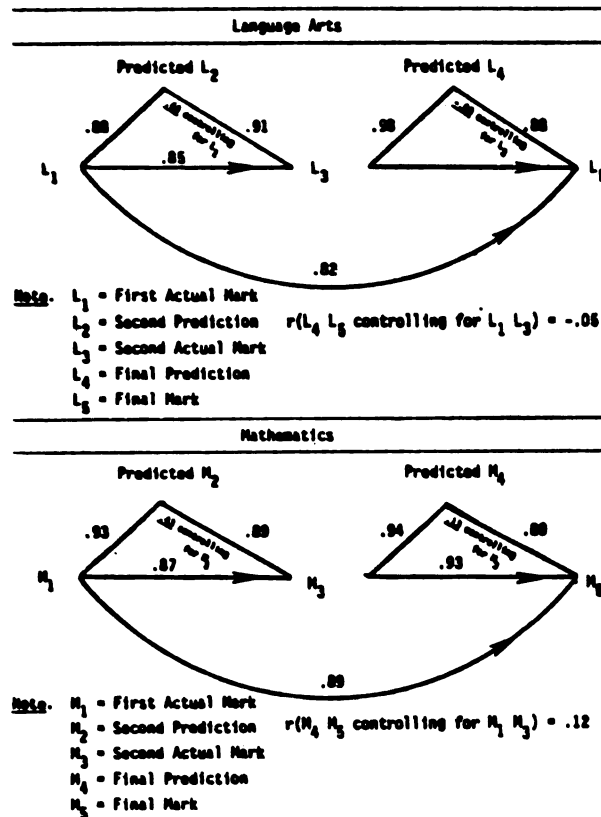
1. If a mark is half way between, I look back in my book to see minuses and pluses which have been given at work periods, i.e., experiments in science. If they have a lot of minuses because of fooling around in class, then I would give them the lower of two grades.
2. The time of year is also important because of motivation. At the beginning of the year, I would tend to mark down. I think that if they start off with a top grade the first marking period, they tend to go down because they don't really have anything to work for. At the semester break, however, I tend to mark up. I feel they ought to have credit for what they have done.
3. Marks are related to motivation, but they are limited by ability. If they get a B, I think that's as far as they can go. Even if they are motivated to try harder, it's pretty hard to bring it up to an A if you don't have that something extra.
4. Marks are especially related to motivation depending on the family situation. It all depends what the parents' expectations are. In most cases, parents are motivated by their kid's marks. Though I have some that really don't care. That is, they care, but it doesn't really make them or get them to force their kids to do a little bit more work or try a little harder. They tend to give up and say, well, that's all their child can do.
5. I have hardly any Es because I don't believe in giving Es. If a child hands in his work, I feel I can't give him an E. I usually push the kid hard to get the work done. Actually, I did give one E because I just couldn't budge him even keeping him in the classroom during recess.

6. No, I don't believe B is an average mark. I didn't realize my marks averaged B. I don't know how this happened because I would say that I have as many outstanding as low students and all the rest in the middle. I'm surprised.

In line with these rules, Teacher Four set up a record book system. She also set up a card system for mathematics where the students are moved from one group to another upon mastery, hence, the record book is too limited. It is the marks derived from this system that are analyzed statistically in the next section. The interview concerning the marks is analyzed following that.

Statistical Analysis

Three statistical techniques (multiple regression, Pearson correlations and partial correlations) were combined to reveal the marking policies of Teacher Four across a school year in language and mathematics. The tables used to modify the original regression equation and yield on adjusted marking model may be found in the Appendix D.

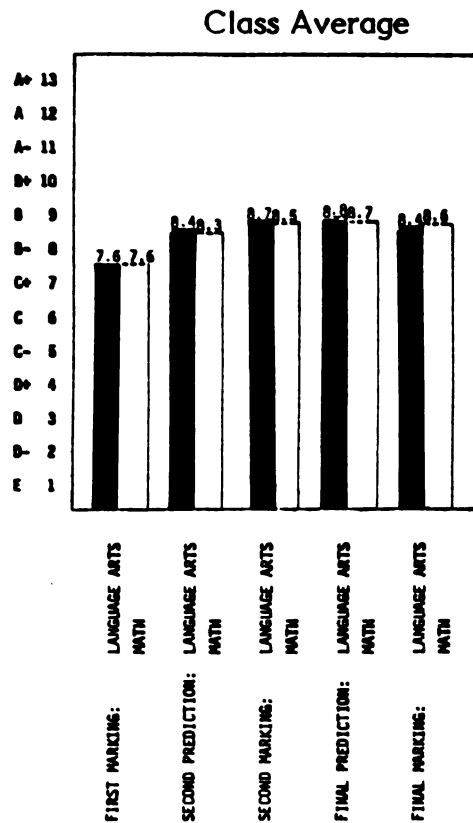


Note. These policies were captured through Pearson correlations adjusted by partial correlations. Summative marks and predicted marks were the base data.

Figure 4.11. Marking policy for Teacher Four (with predictions) for 33 students.

The initial regression equation indicated that the second language mark (Beta = .93) was the best predictor of the final language mark. In mathematics, the best predictor was also the second mark (Beta = .38 and .89 in Stepwise regression) and the only one with any significance. This was supported by Pearson correlations and the partials. It was further confirmed by the frequency bargraph which shows that the second marking average is closer to the final average than the prediction, however, the distinction is minor when the mark averages are so close.

Table 4.21
Pattern of Average Marks Across a Year
Teacher Four
Students 33



Note. These averages were derived from the marks and predicted marks of 33 students across a school year.

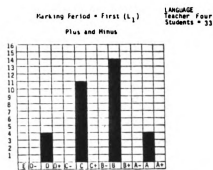
Teacher Four's policies resemble the composite model. The conclusion appears that her marks in general are more highly correlated to actual marks than to predictions. Predictions are more highly correlated to immediate past marks than to future marks. These correlations support the averaging of grades across the year.

A frequency distribution table depicts the marks. The pattern of marks across the year for Teacher Four indicates that she, like Teacher Three, gave a significant group of Ds in the first math mark. Those Ds were displaced by Cs the second marking and by Bs in the last. Like Teacher Three, Teacher Four did not

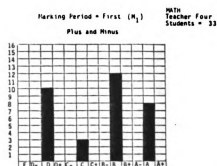
use pluses or minuses on the report card. Her categories were clean. However, Teacher Four did use plus and minus concepts in the record book indicating contingency factors at other levels as shown in Table 4.22.

An analysis of Teacher Four's record book indicates the task focus (see Figure 4.12). Of nineteen entries within one marking, eighteen were corrected and one checked in (- or +). Of these entries two were tests, and the rest were concerned with writing mechanics and parts of speech. The assigned tasks were skill based, hence, less contingency bound, which supports the clean categories of mark distributions.

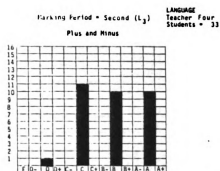
Table 4.22
Distribution of Marks Across Three Marking Periods



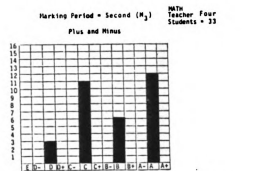
Note: % of minuses & pluses = 0
% of minuses = 0
% of pluses = 0
% related to A/B = 0
Predefined categories = 100%



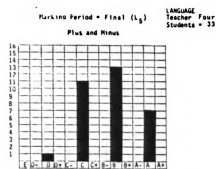
Note: % of minuses & pluses = 0
% of minuses = 0
% of pluses = 0
% related to A/B = 0
Predefined categories = 100%



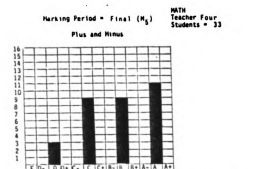
Note: % of minuses & pluses = 0
% of minuses = 0
% of pluses = 0
% related to A/B = 0
Predefined categories = 100%
Blank = 1



Note: % of minuses & pluses = 0
% of minuses = 0
% of pluses = 0
% related to A/B = 0
Predefined categories = 100%
Blank = 1



Note: % of minuses & pluses = 0
% of minuses = 0
% of pluses = 0
% related to A/B = 0
Predefined categories = 100%
Blank = 1



Note: % of minuses & pluses = 0
% of minuses = 0
% of pluses = 0
% related to A/B = 0
Predefined categories = 100%
Blank = 1

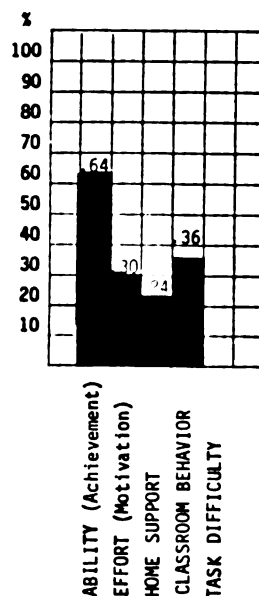
Verbal Analysis

Teacher Four was succinct but clear in her responses to the interview questions and in her commentary about her students. Hence although her comments indicated great care of her students, there was very little elaboration or anecdote from which to extract classroom concerns.

In the attribution-utility comments, Teacher Four stressed ability (.64) followed by classroom behavior (.36). Effort ranks lower than class behavior (.30), however, Teacher Four commented elsewhere in the protocol quite extensively about effort reflecting a limit on effort at both the top and bottom of the scale. Teacher Four clearly stated that effort without ability cannot get an A, and at the other end of the scale almost any effort will merit something above E.

Table 4.23

Attribution-Utility Percentage Teacher Four Students 33



Although her attribution-utility internal pattern differed from the composite model, Teacher Four's comments in the model categories clearly indicated the same judgment cues as other subjects for the marking decisions.

Teachers Four and Three both stressed a skills orientation, which was hypothesized to keep students on task for a grade exchange. Teacher Four, however, mentioned behavior and discipline considerably more. She also displayed fewer As although she gave more Bs. This may speak to the power of the performance-grade exchange to keep students on task. It may also speak to a large class or to the particular student makeup within it. Assuredly it weakens the argument that a clear skills orientation without other factors present can control on-task behavior. From the interview data it is difficult to draw any conclusions beyond the attributional-utility coded comments. Teacher Four stated at one point that she did not consult student records unless she had a specific problem (see rules), but that single statement does not allow much inference. However, Teacher Four voiced the strongest objection to the simulation situation in which 30% of a given class's grades were Cs and Ds at the end of the year. She felt it was likely that the teacher had not been doing the job. Teacher Four used short-term tasks and units as feedback on the learning process long before the summative mark was averaged.

Distinctive to Teachers Four and Two was a clear statement on the impact of time of year on marking. This became a contingency rule and her pattern of average marks indicated an increased class average in January which sets a stable average for the final mark.

Lastly, Teacher Four resembles the composite case in her optimism about the students. Her predicted and actual marks are high. Her cue sort at the end of the year is noteworthy, with 31 students having average to high ability and only eight having low effort.

Table 4.24

Cross Tabulations of Effort and Ability
Teacher Four

		EFFORT		
		High	Low	
ABILITY	High	23	8	31
	Low	1	0	1
		24	8	32

Summary

The marking policies of Teacher Four resemble the composite model. She emphasized skill assignment and marked overwhelmingly by procedural rules. In a few contingency situations, Teacher Four resembled the composite model in her attribution-utility cues with ability and effort being predominant. Teacher Four's judgment process is coherent, task focused and highly classroom bound.

Case Five - Teacher Five

Base Data

Sex: Male

Years of teaching: 17. In this district: 17.

Class size: 31.

Grade and composite: Sixth; 19 girls, 12 boys.

Parents attending last conference: 31 of 31.

Philosophy and Rules

"If I had my druthers, I wouldn't have grades. But society demands them, and kids are not used to not having them. When I used a completely individualized math program with pluses indicating skills mastered, I've seen this cause students to relax and not try as hard as they could. So I suppose we are stuck with this A B C D thing. I would prefer anecdotal records." Teacher Five feels the current marking system which allows comments and encourages conferences along with A B C grades is satisfactory under the circumstances, however, he would prefer more marking periods than four (sample report card in Appendix B). "In all fairness to parents, unless we make phone calls ahead, boom, several weeks go by before they realize their child is going down or not producing. Then what can they do?"

Teacher Five saw marks as communicators of achievement to the home, official records and motivators. To this end, he set up certain rules.

Procedural rules:

1. I mark each subject differently with a range of 5 to 40 items in a particular subject. The items represent homework, skills mastered, essays and tests. I do not mark discussions.

2. All items are planned ahead, however, whether or not they are actually recorded in the book depends on the general success of the lesson, the amount of interruptions I've had, the mood and attentiveness of the students. I'm real flexible. In fact, I keep repeating things in my lesson plans, and I think the principal wonders what's going on.
3. I average grades strictly and then when I see the average comes out between two marks, I take into consideration effort, but not until. (4 points for an A, 3 for a B, etc.) I do it in my head.
4. Parents are concerned with Cs. I do an awful lot of work with parents, send home weekly reports, so parents are right on top of it. And kids know we care and that we are communicating with their parents, and I think they strive better.
5. Weighting. In the individualized math program, I don't grade their daily papers A B C D because in lots of cases, they are learning new concepts and so I only grade their tests. But they do receive a check for doing the work, a minus if they do less and a check plus if they do more. I consider these as effort.

In social studies, I weight tests more than other work. In language, I find marking hardest. If its cut and dried, true or false or picking out details, that's easy. But if it's interpretation or critical thinking, then I think that is the toughest to mark. I pretty well average language out whether it be homework, whether it be a creative writing thing that they've done. Testing I don't do too much of in language.

Contingency rules:

1. If a mark is between two marks and all written work is in, if the child does a lot of class participation and is really trying, I will give that child the higher mark. Where if there is little participation and they talk to their neighbor, I'll go down.
2. I also take into consideration split homes and a child's self-confidence. I use minuses and pluses to help with self-confidence.
3. Definitely, I think many low grades put the child down and that doesn't help their self-concept one bit. I especially do not like giving Es and will do everything I possibly can to not give a child an E. Ds don't help a self-concept either, but they are the only way to show that kids are not producing at all or up to a reasonable expectation. I did give two Es strictly because these kids goofed off the whole year, and I don't want them to get lost in the shuffle of junior high. They need to be followed up.

to the \mathcal{H}_1 and \mathcal{H}_2 are \mathcal{H}_1 and \mathcal{H}_2 respectively.

Let \mathcal{H}_1 and \mathcal{H}_2 be two subspaces of \mathcal{H} such that $\mathcal{H}_1 \cap \mathcal{H}_2 = \{0\}$ and $\mathcal{H}_1 + \mathcal{H}_2 = \mathcal{H}$. Then \mathcal{H}_1 and \mathcal{H}_2 are called complementary subspaces of \mathcal{H} .

Let \mathcal{H}_1 and \mathcal{H}_2 be two subspaces of \mathcal{H} such that $\mathcal{H}_1 \cap \mathcal{H}_2 = \{0\}$ and $\mathcal{H}_1 + \mathcal{H}_2 = \mathcal{H}$. Then \mathcal{H}_1 and \mathcal{H}_2 are called complementary subspaces of \mathcal{H} .

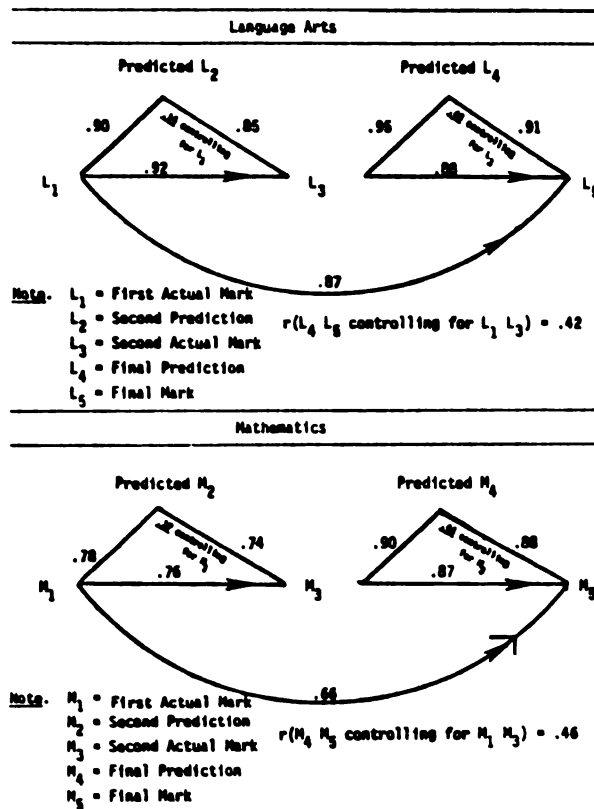
Let \mathcal{H}_1 and \mathcal{H}_2 be two subspaces of \mathcal{H} such that $\mathcal{H}_1 \cap \mathcal{H}_2 = \{0\}$ and $\mathcal{H}_1 + \mathcal{H}_2 = \mathcal{H}$. Then \mathcal{H}_1 and \mathcal{H}_2 are called complementary subspaces of \mathcal{H} .

Let \mathcal{H}_1 and \mathcal{H}_2 be two subspaces of \mathcal{H} such that $\mathcal{H}_1 \cap \mathcal{H}_2 = \{0\}$ and $\mathcal{H}_1 + \mathcal{H}_2 = \mathcal{H}$. Then \mathcal{H}_1 and \mathcal{H}_2 are called complementary subspaces of \mathcal{H} .

4. I have found over the years that the last marking period kids have a tendency to slough off. So I get tough on them the third marking, and they tend not to relax too much in their effort. I would deliberately be harder on the third marking.

Statistical Analysis

Three statistical techniques (multiple regression, Pearson correlations and partial correlations) were used to analyze Teacher Five's marks for 31 students. The tables used to modify the original regression equation and yield an adjusted marking model may be found in the Appendix D. The marking policies for language and math across a school year follow:

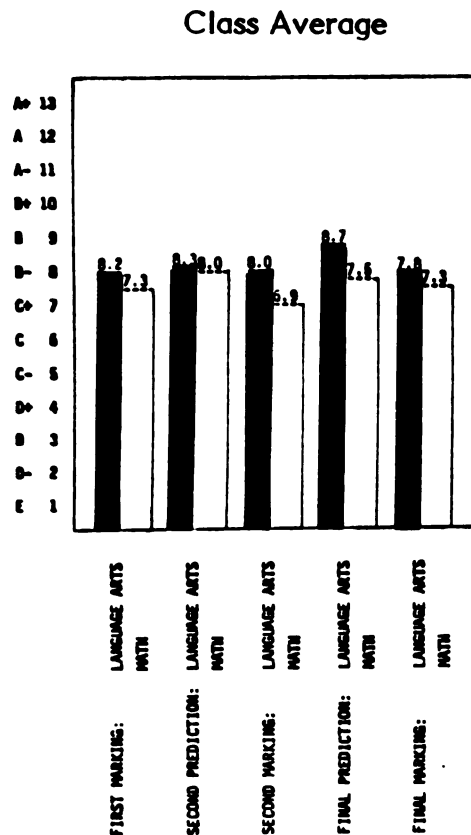


Note. These policies were captured through Pearson correlations adjusted by partial correlations. Summative marks and predicted marks were the base data.

Figure 4.13 Marking policy for Teacher Five (with predictions) for 31 students.

The marking policies resembled the composite case. All marks in Case Five correlated highly and significantly. There was general consistency between language and math. The regression equation was modified by the partial correlations, and that adjustment appeared to be in line with the frequency bargraph which indicated that the class average prediction in language was almost a point higher than the actual mark whereas the actual second marking was only two off.

Table 4.25
Pattern of Average Marks Across a Year
Teacher Five
Students = 31



Note. These averages were derived from the marks and predicted marks of 31 students across a school year.

The frequency distribution of marks for Teacher Five is depicted in Table 4.26.

The criteria of marking for Teacher Five was related to task completion at a given standard (corrected papers or mastered skills). This was averaged across the record period, although less "strictly" than he perceived. Teacher Five followed procedural rules in the majority of cases; corrected tasks according to a standard, weighted tasks in each subject and averaged strictly. An analysis of Teacher Five's record book indicated his marking basis. The first marking in language included eighteen entries. Only two of these was a check-in (✓)

assignment. Two other columns indicated checks for extra work. Eight assignments were written poems, one was a spoken book report, four were in the language textbook Anchors Away, one was an essay. There were no tests in language during this period. Reading was a separate subject as was spelling and penmanship. The first marking grade was an average of the corrected papers and assignments (see Figure 4.14).

Verbal Analysis

"Then when I see the average comes out between two marks, I take into consideration effort, but not until." Then Teacher Five followed contingency rules based primarily on effort and self-confidence. "I'll give grades maybe to help their self-confidence, rather than give them the low grade, I'll put the minus in

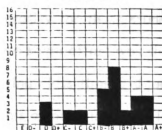
like a C- vs. a D. I'll do that." Teacher Five appeared to have two categories of effort in his attribution system: one category was absorbed into the completed task symbolized in the record book; another category involved extra effort factors such as "does a lot of class participation," "is really trying," "doesn't talk to their neighbor." Hence, although the attribution coding schema allowed only one check per category, it appeared to be a weighted check in the judgment of Teacher Five.

Essentially, Teacher Five resembled the Composite Case in most procedural features, but differed in the number of contingencies and in a lower class average mark (final average equaled C+). Teacher Five's marks were consistent with his statements about contingency situations. His marking pattern over three marking periods bore this out. Thirty-one percent of his marks were pluses or minuses in the first marking, fifty-eight in the second and thirty-four in the final marking. This was consistent with Teacher Five's statements in the protocol that the middle marking was often lower in order that kids don't slough off. It is also consistent with his bargraph which showed that all marks were somewhat lower at the middle marking. Teacher Five differed from the composite in the minuses and pluses he gave, but he represented the general pattern of the composite in having less minuses and pluses in the first marking (see Table 4.26).

Table 4.26
Distribution of Marks Across Three Marking Periods

Marking Period = First (L_1)
LANGUAGE
Teacher Five
Students = 31

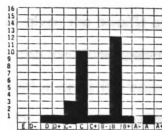
Plus and Minus



Note: % of minuses & pluses = 31.0 % related to A/B = 25.5
% of minuses = 25.5 Preordained categories = 68.0
% of pluses = 6.5 Incomplete = 1

Marking Period = First (M_1)
MATH
Teacher Five
Students = 31

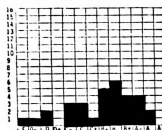
Plus and Minus



Note: % of minuses & pluses = 22.5 % related to A/B = 6.4
% of minuses = 12.9 Preordained categories = 77.5
% of pluses = 9.6

Marking Period = Second (L_2)
LANGUAGE
Teacher Five
Students = 31

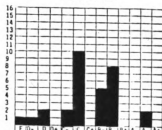
Plus and Minus



Note: % of minuses & pluses = 58.0 % related to A/B = 41.9
% of minuses = 41.9 Preordained categories = 42.0
% of pluses = 16.1

Marking Period = Second (M_2)
MATH
Teacher Five
Students = 31

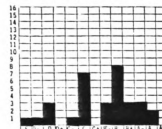
Plus and Minus



Note: % of minuses & pluses = 25.8 % related to A/B = 16.1
% of minuses = 25.8 Preordained categories = 74.2
% of pluses = 0

Marking Period = Final (L_3)
LANGUAGE
Teacher Five
Students = 31

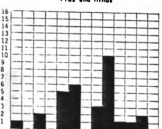
Plus and Minus



Note: % of minuses & pluses = 34.5 % related to A/B = 29.1
% of minuses = 24.5 Preordained categories = 65.9
% of pluses = 9.7

Marking Period = Final (M_3)
MATH
Teacher Five
Students = 31

Plus and Minus

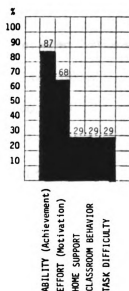


Note: % of minuses & pluses = 32.2 % related to A/B = 16.1
% of minuses = 19.0 Preordained categories = 67.8
% of pluses = 1.2

Teacher Five's attributional categories resembled the general pattern of the composite case with ability leading effort and with significant comments on home support and classroom behavior.

Teacher Five discussed home support problems in detail, expressing sympathy with many students' situations. However, his marks did not show extra generosity for those students. Students were given the benefit of the doubt only where some effort was shown. Teacher Five also mentioned puberty as a cause of wasting class time and lack of work. It was not a cause for leniency, but neither did his comments reveal inordinate concern. He has the sixth grade and portrayed puberty problems of growth and "not getting his act together" as characteristic of some sixth graders yearly. Although Teacher Five was concerned about a student's home support and classroom behavior, he modified the impact of these judgmental cues by mediating them through the cues of effort, i.e., checks, check minuses and check pluses in his record book. Effort adjustments were only a minus or plus away from the original calculated average.

Table 4.27
Attribution-Utility Percentage
Teacher Five
Students 31



Teacher Five differed from the attributional-utilities composite in his interest in task difficulty. He was as concerned with that category as he was with home support and classroom behavior. He was especially interested in a solution for students with test blocks which he categorized as a task difficulty.

Teacher Five was optimistic about his students as shown in the cross tabulations of effort and ability where 26 of 31 pupils are considered to have average to high ability and 22 had high effort. Although not statistically meaningful, the crosstabs gave description to the class from the teacher's perspective

Table 4.28
Cross Tabulations of Effort and Ability
Teacher Five

		EFFORT		
		High	Low	
ABILITY	High	21	5	26
	Low	1	4	5
		22	9	31

There was consistency between optimism, high predictions and Teacher Five's marks. He gave 19 marks B and above.

Consistent with his attitude toward effort, low marks and confidence, Teacher Five gave only one E in language and one in math for complete "goofing off" all year. He also gave a D- to avoid an E.

Summary

The judgment process of Teacher Five supports the model. In the majority of marks, he followed procedural rules. In a minority, he used contingency rules.

which Teacher Five saw as highly related to self-confidence. Physical maturity was a significant part of the class behavior category, six comments out of nine on puberty or size.

The marking process of Teacher Five was primarily bounded by the classroom. He was interested in parent influence to help with self-confidence and effort, and self-confidence was directly related back to effort. Teacher Five, however, was one of two teachers to mention the next grade and the need for two students to be followed up in junior high school. He mentioned an interest in the California Basic test results to let him know if his class expectations were realistic. This interest in outside factors may be a function of the sixth grade which leads to the next school.

In summary, Teacher Five evidenced a coherent marking policy which supported the model and which was focused on and bounded by classroom tasks.

Case Summary

The five teacher cases reveal marking policies and strategies which corroborate the composite case. For most marks, teachers follow the procedural rules. One teacher in this study differed significantly from the composite by using contingency rules more than procedural rules.

The teachers agreed upon the five important cues operating in contingency situations; ability, effort, home support, classroom behavior and task difficulty. However, each teacher weighted these cues differently as can be seen in the attributional coding tables.

At the level of teacher case analysis, some additional hypotheses emerged involving interrelationships between contingency judgment cues. These appear to involve trade-offs. Within this study, there was no systematic way to code for these hypotheses, but they need to be explored in future research.

- The pattern of ascendance and recession of certain cue categories appears to be related to the fact that effort must be sustained over 180 days during a year's period. Such intensity of social conditions may demand compromises between achievement factors and behavior factors, hence, the ascendance of class behavior and home support during the midyear.
- The influence of home support level appears to be related to the time of year with home support being especially important between the first marking/conference and the next to the last marking/conference. Apparently, teachers use the first marking period as a time of assessment of factors in existence. By the last marking period, they are not only aware of the categories but they are aware of the extent to which such categories may be combined to bring about task completion.
- The category of classroom behavior and physical maturity also appears to be related to time of year. It appears to have greatest impact on teacher judgment at the beginning of the year, descending slightly after the first half of the year when class routines have been established.

- The categories of ability and effort are dominant throughout the year, however, they too seem to be related to a time pattern, with almost overwhelming influence during the first and last marking period.
- The importance of task difficulty as a cue category for marking appears related to the extent to which it is a consideration during planning and task selection. Where tasks are originally assigned at an appropriate level of challenge, task difficulty recedes as a marking judgment cue. Grade level reading materials, individualized programs and special education prescriptions relieve teachers of one complex judgment process. When tests were failed by a significant number of students, teachers tended to throw-out results and reteach rather than use a "task difficulty" category.

These hypotheses are very tentative, but they have some support in the frequency distribution tables which indicate patterned fluctuations across the year. They need consideration in future research.

The teacher cases, taken together, strengthen the composite model of the marking judgment process.

Integrated Summary

This chapter displayed and analyzed the judgment, policies and cues of five teachers during the marking process across the school year. Evidence was presented in a composite case and five teacher cases. A model of the judgment process was constructed from the rules and cues which emerged during investigation.

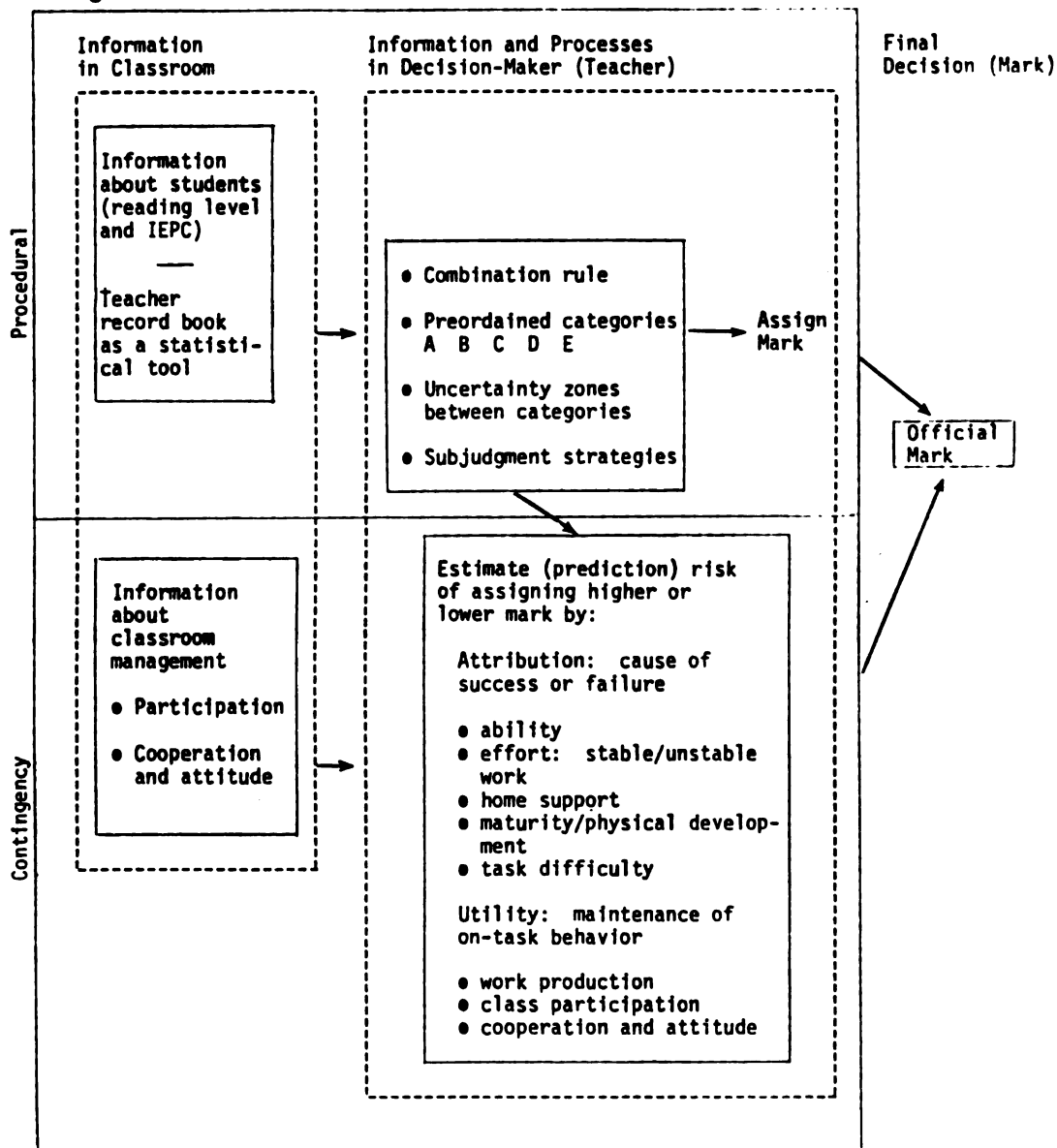


Figure 4.15. Framework for marking process (adapted from Carroll and Payne, 1976).

The model indicates that teachers generally follow procedural and contingency rules which divide the process into three stages: collection of task completion information, computation of task information and modification strategies to deal with uncertainty between categories and with failure.

In the majority of 152 cases, teachers marked pupils routinely from record book information which was combined according to a preordained category of A B C D or E (F). This was a linear process operating across the top of the model according to procedural rules. The primary cue used in the marking at the procedural rule level was task completion at a given standard (formative mark) or completion as a check (✓) in or out.

In a minority of 152 cases, the combination rule resulted in uncertainty between two categories or in failure. Then contingency rules were put into operation. Contingency rules involved attributional and utility strategies based on consideration of five factors or cues which emerged from verbal analysis: effort, ability, home support level, classroom behavior and physical maturity and task difficulty. Prediction was not a judgment cue in itself. Under contingency rules, effort appeared as the primary cue vying with ability.

Elementary teachers did not use summative marks as a feedback mechanism to improve teaching. Instead, they used intermittent tasks such as tests. If a significant group of students failed, the teachers judged themselves unsuccessful in teaching the unit and, therefore, retaught or discarded the grade. The expectation that summative marks should serve in a feedback capacity is misplaced.

The major conclusion of the study is that teachers have a coherent marking judgment process which operates across a school year. Within this process, task completion at a stated level of difficulty and at a given standard of mastery is the dominant cue in the marking judgment. Other cues operate in

zones of uncertainty between two preordained marks or in exceptions such as failure. The judgment process is bounded by the classroom task environment.

CHAPTER V

CONCLUSIONS AND IMPLICATIONS

Introduction

The persistent dissatisfaction with traditional marks, A B C D E, which symbolize pupil progress, prompted this investigation of teacher marking processes. A review of the voluminous research literature on marks and the emerging literature on teacher decision making revealed little systematic inquiry into the judgment process underlying marks. The purpose of this study was to develop an understanding of the marking process which engages teachers across a school year by describing the heuristics, strategies and cues of five upper elementary teachers (152 students) from a typically achieving school district in Michigan. The study posed seven research questions and investigated them through established research methods from the field of human judgment: process tracing, policy capturing, utility theory and attributional theory. Chapter V summarizes the findings under the question headings, compares them to the functions ascribed to marks by society (Chapter I), and discusses the implications for practice and future research.

Summary of Findings by Research Questions

Upon what information is the summative mark based? The summative mark for each marking period is based upon the completion of a significant number and variety of assigned tasks at an appropriate level of difficulty and standard of mastery.

What cognitive processes make possible the formative stages (record book categories) of marking? The cognitive processes of selection, simplification and inference operate through heuristics (rules), attributions of individual success and failure and perceived utilities of the classroom. Procedural rules emerged which guide and routinize the marking process. The record book is the key inferential tool of the process. Each teacher had variations on these rules, but all specified a significant number of tasks, a variety of tasks and an appropriate level of difficulty. The specification of tasks rested on the basic assumption that student learning results from completing meaningful tasks.

Is there a judgmental rule which explains how the input information (formative) is transformed into the output (summative) mark? There is a linear arithmetic rule which averages across collected marks and which is directly related to standard of mastery and degree of task completion. Within a marking period, this rule is focused on completed tasks which carry weighted values and are assigned to the preordained categories of A B C D E. For example, ten math points earn an A, nine a B, and so on. In turn, each A is worth 4 points, each B is worth 3, each C is worth 2, each D is worth 1. There is a great discrepancy as to whether an E equals 0 or something above 0. Across the year, the rule focuses on averaging the summative marks of each marking period. Hence the final mark is a derived arithmetic mean based on the weighted values of the completed tasks of each marking period.

If the judgmental rule yields a zone of uncertainty between any two preordained categories or yields a failure, what cognitive processes enable the teacher to mark up or down? Whereas procedural rules emerged to organize the marking process, contingency rules emerged to help clarify choices in uncertainty. Contingency rules rested on attributions of individual student success or failure and perceived utilities for total classroom behavior. Attribution

and perceived utility are inferential thinking processes which go beyond the collected data. In this study, they were encompassed within the categories of ability, effort, home support, classroom behavior/physical maturity and task difficulty. The most common tools for assessing these attributions or utilities were checks, minuses and pluses.

Other conditions influenced contingency judgments. These included (1) trade-offs between contingency categories, (2) time of the 180 day-year and (3) extreme absence without cause. Systematic inquiry into these conditions was not within the scope of this study.

Do identified cognitive processes form a pattern, schema or model of the marking process? A model has been proposed. This model is based on the procedural and contingency rules which divide the marking process into three phases: selection and collection of data, valuing and assigning of data to preordained categories of A B C D E, and contingency factors to facilitate choice under uncertainty or failure. The majority of marks are determined at the procedural level.

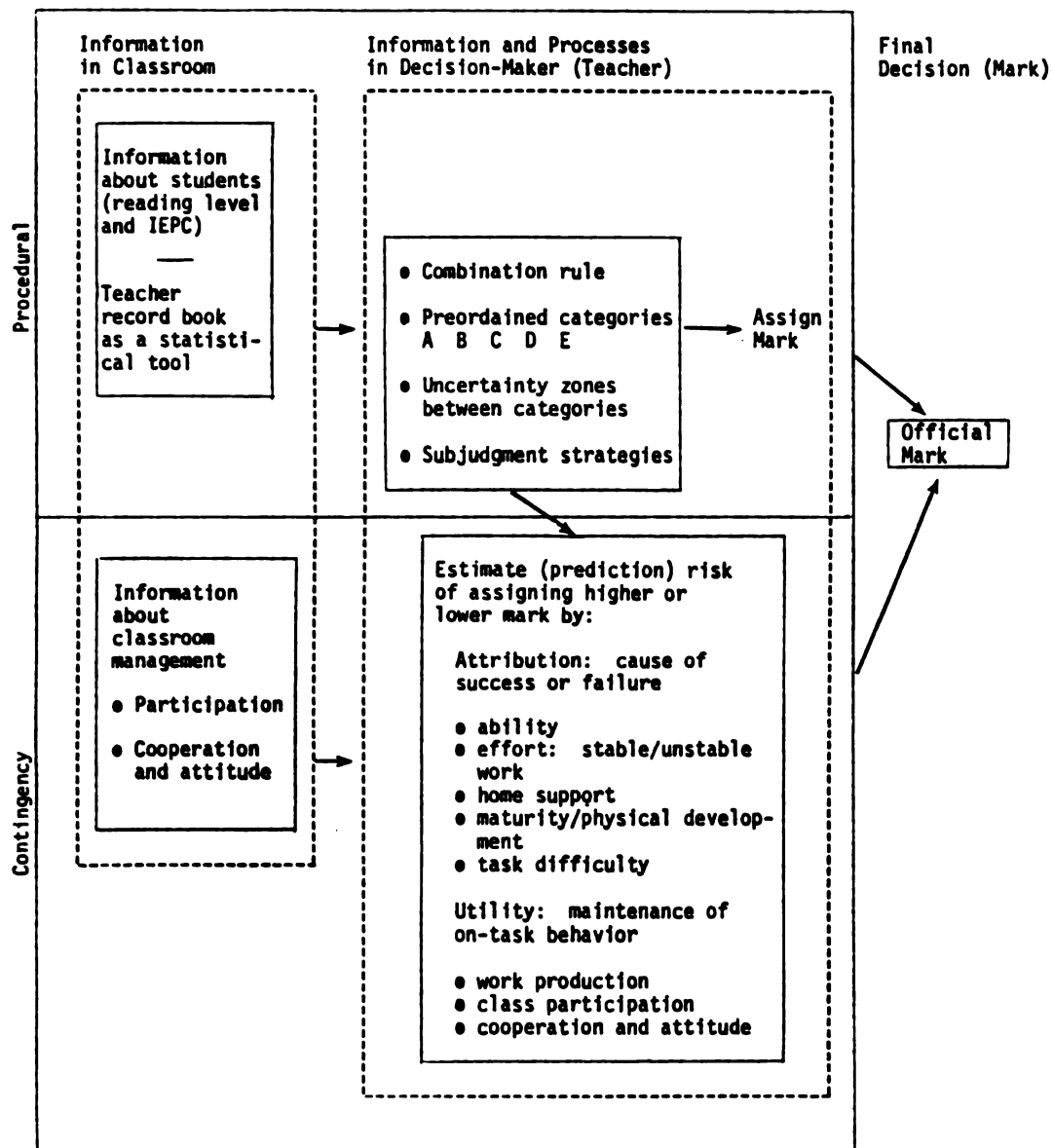


Figure 5.1 Framework for marking process (adapted from Carroll & Payne, 1976).

Do identified cognitive processes account for the five functions ascribed to marks by society in general? A review of the ascribed functions which were presented in Chapter I, indicates that the functions may be classified into two general groups: one involves assumptions about marks related to conditions outside the classroom such as future counseling placement within the K-12 program, future marks, and future job success; the other involves conditions inside the classroom (ecology) structure such as motivation, achievement and a teaching feedback function. The findings of this study indicate clearly that the judgment

processes of teachers (rules, strategies and cues) are focused on task completion which is bounded by the particular classroom and its immediate participants. The marking judgment processes of teachers, therefore, are not concerned about the functions ascribed to marks which are outside the classroom. Marking judgments primarily relate to task completion at a given level of difficulty and standard of mastery, and to the factors which promote that completion. Hence teachers define their marking responsibility in terms of the practical demands of 30 pupils in a classroom for a whole year.

Of the four methods of investigation used, is one superior for illuminating the marking process? The four methods of process tracing, policy capturing, attribution theory and utility theory, shed light on different levels of the marking model. Process tracing provided the broadest description of the marking judgment and supplied some part of the answer for each research question. It provided the most rules and cues used in the marking process across the year. Consistent with the process tracing discussion by Einhorn in Chapter III, the distinction between two subjudgment phases emerged; one which dealt with choices between multiple categories, A B C D E, and one which dealt primarily with a choice between any two categories. These phases were labeled procedural and contingency, and they provided the major divisions of the model. A definite weakness of process tracing was its inability to distinguish the various weights of factors in the judgment.

Policy capturing dealt best with the procedural questions, with the summative marks across the year and with teacher choices between multiple categories of marks. It answered research questions pertaining to combination rules across the year leading to the conclusion that each marking period functions separately. Within policy capturing, different statistical techniques led to different results. For example, multiple regression tended toward a recency effect unless adjusted. Pearson correlations made a repeatedly strong case for a

primacy effect. Partial correlations tended to adjust both techniques and supply a modified policy which led to a neutral position on recency and primacy effects. This neutrality forced attention back to the significance of formative marks within the record book.

Attribution theory dealt well with the research questions regarding zones of uncertainty between any two categories. Protocol comments were categorized and counted illustrating the general weighting of the categories of ability, effort, home support and task difficulty. Adjusted attribution charts showed effort to predominate the judgment but always vying with ability. This substantiated Weiner's findings discussed in Chapter III, page 63. Policy capturing with statistical analysis did not get at these factors, but attribution theory with verbal analysis did. The findings were further supported by frequency distributions of pluses, minuses and checks, which showed that contingency situations tended to increase as the year progressed. Attribution theory, however, is oriented toward an individual psychology, and it misses some aspects of cooperative class behavior.

Utility theory filled in the class behavior gap. It too was concerned with contingency factors particularly on-task behavior. Some teachers gave pluses and minuses in separate columns specifically for cooperative behavior. These columns were only consulted when a mark was determined to be in a zone of uncertainty. The decision tree tool illustrates teacher risks and thoughts when deciding to give a higher or lower grade. Utility theory is very concerned with estimating future effort or behavior, but not concerned with attributing cause on an individual basis.

It can be concluded that the research question was inappropriately phrased when it asked for a superior method. Instead, each method had strengths and weaknesses. Together they illustrated the total, year long marking process with its emphasis on task completion. The four methods together led to the identification of a model of the cognitive processes involved in marking

1997-1998, 2000-2001, 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013, 2014-2015, 2016-2017, 2018-2019, 2020-2021, 2022-2023, 2024-2025, 2026-2027, 2028-2029, 2030-2031, 2032-2033, 2034-2035, 2036-2037, 2038-2039, 2040-2041, 2042-2043, 2044-2045, 2046-2047, 2048-2049, 2050-2051, 2052-2053, 2054-2055, 2056-2057, 2058-2059, 2060-2061, 2062-2063, 2064-2065, 2066-2067, 2068-2069, 2070-2071, 2072-2073, 2074-2075, 2076-2077, 2078-2079, 2080-2081, 2082-2083, 2084-2085, 2086-2087, 2088-2089, 2090-2091, 2092-2093, 2094-2095, 2096-2097, 2098-2099, 2100-2101, 2102-2103, 2104-2105, 2106-2107, 2108-2109, 2110-2111, 2112-2113, 2114-2115, 2116-2117, 2118-2119, 2120-2121, 2122-2123, 2124-2125, 2126-2127, 2128-2129, 2130-2131, 2132-2133, 2134-2135, 2136-2137, 2138-2139, 2140-2141, 2142-2143, 2144-2145, 2146-2147, 2148-2149, 2150-2151, 2152-2153, 2154-2155, 2156-2157, 2158-2159, 2160-2161, 2162-2163, 2164-2165, 2166-2167, 2168-2169, 2170-2171, 2172-2173, 2174-2175, 2176-2177, 2178-2179, 2180-2181, 2182-2183, 2184-2185, 2186-2187, 2188-2189, 2190-2191, 2192-2193, 2194-2195, 2196-2197, 2198-2199, 2200-2201, 2202-2203, 2204-2205, 2206-2207, 2208-2209, 2210-2211, 2212-2213, 2214-2215, 2216-2217, 2218-2219, 2220-2221, 2222-2223, 2224-2225, 2226-2227, 2228-2229, 2230-2231, 2232-2233, 2234-2235, 2236-2237, 2238-2239, 2240-2241, 2242-2243, 2244-2245, 2246-2247, 2248-2249, 2250-2251, 2252-2253, 2254-2255, 2256-2257, 2258-2259, 2260-2261, 2262-2263, 2264-2265, 2266-2267, 2268-2269, 2270-2271, 2272-2273, 2274-2275, 2276-2277, 2278-2279, 2280-2281, 2282-2283, 2284-2285, 2286-2287, 2288-2289, 2290-2291, 2292-2293, 2294-2295, 2296-2297, 2298-2299, 2300-2301, 2302-2303, 2304-2305, 2306-2307, 2308-2309, 2310-2311, 2312-2313, 2314-2315, 2316-2317, 2318-2319, 2320-2321, 2322-2323, 2324-2325, 2326-2327, 2328-2329, 2330-2331, 2332-2333, 2334-2335, 2336-2337, 2338-2339, 2340-2341, 2342-2343, 2344-2345, 2346-2347, 2348-2349, 2350-2351, 2352-2353, 2354-2355, 2356-2357, 2358-2359, 2360-2361, 2362-2363, 2364-2365, 2366-2367, 2368-2369, 2370-2371, 2372-2373, 2374-2375, 2376-2377, 2378-2379, 2380-2381, 2382-2383, 2384-2385, 2386-2387, 2388-2389, 2390-2391, 2392-2393, 2394-2395, 2396-2397, 2398-2399, 2400-2401, 2402-2403, 2404-2405, 2406-2407, 2408-2409, 2410-2411, 2412-2413, 2414-2415, 2416-2417, 2418-2419, 2420-2421, 2422-2423, 2424-2425, 2426-2427, 2428-2429, 2430-2431, 2432-2433, 2434-2435, 2436-2437, 2438-2439, 2440-2441, 2442-2443, 2444-2445, 2446-2447, 2448-2449, 2450-2451, 2452-2453, 2454-2455, 2456-2457, 2458-2459, 2460-2461, 2462-2463, 2464-2465, 2466-2467, 2468-2469, 2470-2471, 2472-2473, 2474-2475, 2476-2477, 2478-2479, 2480-2481, 2482-2483, 2484-2485, 2486-2487, 2488-2489, 2490-2491, 2492-2493, 2494-2495, 2496-2497, 2498-2499, 2500-2501, 2502-2503, 2504-2505, 2506-2507, 2508-2509, 2510-2511, 2512-2513, 2514-2515, 2516-2517, 2518-2519, 2520-2521, 2522-2523, 2524-2525, 2526-2527, 2528-2529, 2530-2531, 2532-2533, 2534-2535, 2536-2537, 2538-2539, 2540-2541, 2542-2543, 2544-2545, 2546-2547, 2548-2549, 2550-2551, 2552-2553, 2554-2555, 2556-2557, 2558-2559, 2560-2561, 2562-2563, 2564-2565, 2566-2567, 2568-2569, 2570-2571, 2572-2573, 2574-2575, 2576-2577, 2578-2579, 2580-2581, 2582-2583, 2584-2585, 2586-2587, 2588-2589, 2590-2591, 2592-2593, 2594-2595, 2596-2597, 2598-2599, 2600-2601, 2602-2603, 2604-2605, 2606-2607, 2608-2609, 2610-2611, 2612-2613, 2614-2615, 2616-2617, 2618-2619, 2620-2621, 2622-2623, 2624-2625, 2626-2627, 2628-2629, 2630-2631, 2632-2633, 2634-2635, 2636-2637, 2638-2639, 2640-2641, 2642-2643, 2644-2645, 2646-2647, 2648-2649, 2650-2651, 2652-2653, 2654-2655, 2656-2657, 2658-2659, 2660-2661, 2662-2663, 2664-2665, 2666-2667, 2668-2669, 2670-2671, 2672-2673, 2674-2675, 2676-2677, 2678-2679, 2680-2681, 2682-2683, 2684-2685, 2686-2687, 2688-2689, 2690-2691, 2692-2693, 2694-2695, 2696-2697, 2698-2699, 2700-2701, 2702-2703, 2704-2705, 2706-2707, 2708-2709, 2710-2711, 2712-2713, 2714-2715, 2716-2717, 2718-2719, 2720-2721, 2722-2723, 2724-2725, 2726-2727, 2728-2729, 2730-2731, 2732-2733, 2734-2735, 2736-2737, 2738-2739, 2740-2741, 27

judgments. Together they answered the research question relating to the five functions of marking by indicating that the validity of past research on marks must be questioned in light of its general limitation to single phases of a much larger judgment process and its general focus on functions outside the classroom. Only with a multimethod approach was the total process illustrated.

Implications for Research

Four outcomes of the study have implications for research: the importance of task completion as the primary unit of the performance-grade exchange; the classroom bounds of the marking process; the value of the multimethod approach to marking judgments; and the heuristic value of the model. These outcomes relate to research in different fields of education.

Task completion at a given level of difficulty and a given standard of mastery emerged as the primary judgment cue of teachers during the marking process. The factor of completion, or the filling in of columns across the teacher record book appears to carry a heavier weight than the quality of the completed work. Two features substantiate this assertion: any work handed in receives some credit above E; students operating at a lower than class average level of task difficulty can receive the same amount of credit. However, it is also notable that above the level of C, teachers begin to create more categories of distinction by the use of minuses and pluses. Note the frequency distribution charts of marks across the year. Hence the criterion of completion has greater weight below C and the criterion quality vies with completion above C. The criterion of completion is greater with students operating below grade level on task difficulty.

This overwhelming emphasis on task completion at both an individual and class level calls into question a prevalent notion that teachers mark students

according to racial or socioeconomic characteristics as implied in some expectancy research. The marking task at the end of a given time period appears to be based on different factors than those used in the prediction process at the beginning of a time period, most notably the factor of completion. The distinction between prediction and judgment has not been clarified in previous studies. The marking judgment relies directly on student task completion and indirectly on the classroom behavior which produces task completion more than it relies on identified student characteristics. This overwhelming emphasis on completion also draws attention to the quality and quantity of the original tasks and the expectations assigned during planning. The current debate about the perceived rigor of private schools (Coleman, 1981) or of the effective public schools (Brookover and Lezotte, 1976) goes to the heart of the issue of assigned and completed tasks. Do teachers assign more tasks at a greater level of difficulty in effective schools? What factors influence the number, variety and quality of assigned tasks? The implication for research is that the teacher expectation studies need to have a student evaluation (marking) dimension.

The second factor which has implications for research is the bounded nature of the classroom. This may explain some of the previous unreliability of marks. The review of the marking literature indicates that most studies compared marks to functions outside the classroom such as future placement and future success. Current studies in teacher decision making and planning, are finding that the classroom culture has its own demands which must be considered. The work of Doyle, in particular, emphasizes the ecological nature of the classroom. The planning studies of Yinger and Clark specifically found that the chief unit of planning was the task rather than behavioral objectives. The implications of this marking study are that future studies of marking must account for the bounded nature of the process. Teacher decision-making research needs to examine the

relationship between tasks and marking, between planning and marking and between time-on-task and weighting of tasks. To date teacher decision-making studies have emphasized the preactive and interactive phases of decision making, neglecting the postactive.

The multimethod approach to marking studies is promising and has implications for research. When tasks have been investigated in the past, only one task such as a test or paper has been examined. For example, the Starch and Elliott model of research asks a significant number of experts (100+) to correct one essay or test and concludes that marks are unreliable. This current study suggests that the reliability of one task is discounted by most elementary teachers who collect a great number and variety of task data in their record books. In the future, research on the number, variety and weighting of assignments promises greater insights than replications of one time task research.

The past habit of examining single products and generalizing the results to the marking process points to the role which the marking judgment model could play. In effect, it provides a framework for past marking studies which shows that some studies were entirely involved with the procedural level of marking, others with the contingency level. Either one alone does not account for the total marking process. Hence, the model places the value of past studies into a meaningful framework.

Implications for Practice

The outcomes of the study have implications for practitioners which are primarily related to the heuristic value of the model. Recalling Stenhouse's earlier concern that the use of research was to map the range of experience rather than to perceive the operation of laws within it, and to work through the

refinement of judgment rather than the refinement of prediction, this marking study adds to his goal. The model can be used as a practitioner tool for reflecting upon aspects of the marking task. Practitioners can ask themselves what data they do collect for a mark. They can examine the quality and variety of their tasks and the extent to which some tasks may represent trivia or depth. They can reflect upon the interrelationships between various contingency factors and upon the relationship between procedural and contingency rules.

The importance of the home support category is cause for reflection. To what extent do practitioners rely upon the home for leverage? To what extent do they communicate their procedural rules to the home rather than being satisfied with the oft repeated combination rule statement that 90 to 100 is an A, 80 to 89 is a B, etc., which is only a very small aspect of marking. In this regard, there are obviously implications for the home. The role of the family in task completion is important and often neglected in discussions of educational accountability. School districts may need to articulate this role to parents and to reexamine the role of homework which many parents actually request.

The fact that many teachers do not use the summative mark at the end of a marking period as a feedback mechanism needs discussion and further exploration. If teachers feel that a variety of tasks are important to reflect a range of student capabilities, then why do they not look at the summative mark which reflects this range as an important source of assessment? Why do they emphasize formative task feedback to the exclusion of summative feedback? There may be important instructional reasons why this is so, but at this time, the problem has not been addressed by practitioners or researchers.

Lastly, there are implications for teacher educators. The model provides the opportunity to discuss the framework for marks and the importance of some consistency between class activities, assigned tasks and weighted marks in the

record book. Rather than leaving the marking process as a last thought after instruction, it needs to be integrated into the entire instructional process. In particular, the potential use of summative marks as an additional source of feedback needs exploration.

Summary

This study was intended to generate a description of the judgment processes of five elementary teachers (152 students) during marking across a school year. The findings support a model of the marking judgment constructed from the strategies and cues which emerged through analysis of marks, record books and interviews. The model presents a three phase process which is guided by procedural and contingency rules. Findings indicate that task completion is the primary focus of the judgment, with the criterion of completion having a variable weight in the judgment. The marking judgment is bounded by the classroom, a conclusion which suggests that many past marking studies have made assumptions about marks which are inappropriate to the teacher judgment process. The study found formative marks serve as a feedback mechanism but that summative and final marks do not. The study was limited to five experienced teachers, hence any specific conclusions are highly tentative. The model, however, is useful as a heuristic to generate further discussion, deliberation and research hypotheses.

APPENDICES

APPENDIX A

INTERVIEWS

The primary means of collecting data was the structured, in-depth interview. Five elementary teachers were interviewed following three different marking periods (November, February, June). The first interview was the longest, taking from one and one-half to two hours. The format of the second and third interview was much shorter as may be seen. However, all interviews asked teachers for each pupil's mark, the predicted mark for the next marking and the reasons for the mark remaining, going up or going down.

INTERVIEW #1 - NOVEMBER

INTERVIEW #2 - FEBRUARY

INTERVIEW #3 - JUNE

Questionnaire

The Grading Process

Interview #1 - November

Introduction

The following questionnaire contains a variety of questions. They are attempt to find out what processes and strategies you use to organize all of the work your students do into a single mark in a subject. At the end of the questionnaire, you may have some thoughts to add, and I would welcome them. In the interest of time, I shall go right ahead with the questions indicating periodically which one we are on for the sake of a quicker review of the tape. Please feel free to comment at the end of the session.

Questionnaire

Section I

1. How many years have you taught?
2. How many years have you taught in this district? this school?
3. Have you ever used a marking system other than A,B,C symbols? (Probe:
-Did you prefer it? -Strenghts and weaknesses?)
4. Has the marking system in this school been changed or examined recently?
5. Are you satisfied with the report card format at this school? (Probe:
-Do you have specific suggestions for change?)
6. How often do you mark cards? Are these times satisfactory? More times?
Less?
7. Does the district give a day off school to complete the marking process?
(Probe: -Do you find this useful? -Do you do any of the marking in
advance?)
8. When you were in elementary school, how did you feel about marks?
9. Do you have a working philosophy about marking and where it fits into
the whole educational picture?
10. Does yourdistrict have a working philosophy about grading? (Probe:
-Do you agree with the symbol system as explained on the report card
or do you have some different ideas.

The following questions relate to your current class and its marks

11. Which grade do you have this year?
12. How many children are in the class? (Probe: -Boys -Girls)
13. On a scale of one to five, to what degree are the ethnic backgrounds of
your students similar. One represents little difference and five is
great difference. (1 2 3 4 5) Question discarded
14. On a scale of one to five, to what extent is the general behavior of the
students similar. One represents litte diversity and five is great.
(1 2 3 4 5) Question discarded
15. On a scale of one to five, to what extent is the achievement of students
similar? One represents little difference or a rather homogeneous class,
five represents great difference. (1 2 3 4 5) Question discarded
16. On a scale of one to five what is the degree of parental involvement in
school activities? (1 2 3 4 5) Question discarded

17. On a scale of one to five what is the degree of parental interest in pupil progress? (1 2 3 4 5) Question discarded
18. What was the number of your students' parents who attended the recent conference.
19. On a scale of one to five, what is the degree of parental stability in the community? i.e. how many one parent families? (1 2 3 4 5) Question discarded
20. How does the achievement level of this class compare to other classes which you have had? (Probe: -comparable, greater, less)
21. Does this class operate at grade level? (Probe: -How many above? -How many below? -How do you determine this? i.e. textbooks, reading level etc.)
22. Can a mark reflect the situation where an individual pupil is progressing but is still below grade level? (Probe: -How?)
23. Can a mark reflect the situation where an individual pupil is achieving much above grade level, but is not working very hard? How?

The following questions are focused around the record book and they are the heart of this study because we really know very little about the ways in which teachers organize the marking task.

24. How many marks are considered in this marking period?
25. What do these marks represent? (Probe: -Subject, tasks -Why were these specific assignments chosen?)
26. Do you plan these categories well in advance or do you wait until the task is over to decide whether it should be in a record book category? Discuss your process
27. Are there activities which occur which you don't mark? (Probe: -Examples)
28. Do you find some subjects easier to mark than others? (Probe: -Which ones? -Why? -How does Math contrast with English?)
29. Could we now look over each students grade at the end of the marking and could you explain to me how you came to the final mark in each case. I am not interested in student names but I am interest in cases where it was a problem deciding which grade to give. i.e. students who fell between two grades. Could you please predict whether each pupil will improve next time, probably remain the same or lost ground?

	Grade	Predicted Grade	Factors Influencing
i.e. student			
29.1 L.A.-			
M. -			

31. Do you think there would be much agreement amongst teachers about various criteria to be considered in marking?
32. Do you feel it is useful to compare the marks which you give with student assessment scores—or other standardized tests? (Probe: a. Have you ever done so? b. Do you record the scores?)
33. Do you feel that marks are related to student motivation? (Probe: a. In your class? How? b. In another class? How? c. Are marks ever a reward?)
34. As you look over this whole group of grades, how would you say this group is progressing? (Probe: -Are you satisfied? -How do you feel about those receiving less than C? -Do you feel that you can help them improve? -Do some need additional resources? Are these available?)
35. You have said that you are generally (satisfied - dissatisfied). Will you change any of your plans for the year? How? (Probe: -Will you use or change groups? -Will you add resources or new activities?)

General Etic validation

36. Have you ever been questioned by parents about a student's mark? (Probe: (1) How often? (2) About what concerns?) How do you handle this situation? (1) Have you ever changed a mark for a parent?
37. Do students question their report card marks? (Probe: (1) How often? (2) What concerns them? (3) How do you handle this situation? (4) Have you ever changed a mark for a pupil?
38. Has the principal ever question report card marks? (Probe: (1) How often? (2) What concerns him? (3) How do you handle this situation?
39. Since I approached you about this study of the grading process, have you had any particular thoughts about the subject in general?

It is possible that as we go through the year, you will have some insights of your own on this topic which would be very important to me. I would like to leave you a small notebook and pen, and if random thoughts occur to you, would you please write them down and I'll pick them up when I come in the new year.

My telephone number is 626-6252, and I would love to discuss this in more detail if you have any questions.

Interview #2 - February

1. In our previous interview, you mentioned that the final grade at the end of a marking period was arrived at by an averaging of marks in the book and not really a difficult decision. When a mark is computed to be half way between two grades, what factors influence your decision to go up or down?
2. You mentioned that you weight tests more heavily than other work. Could you clarify how you weight tests and why they are worth more than daily work?
3. Looking at a particular test, how do you decide which one will merit an A? That is, how do you set the standard for a given test? Before the class takes it or after you have seen results?
4. How would you interpret a test if the majority of the class failed it?
5. Would you discuss your feelings about the relationship between a test score and how much you feel the student has actually learned?
6. What do you feel about the number and timing of tests?
7. You mentioned that marks were a motivating factor. Why do you think this is so?
8. Do marks have the same motivational power at each marking period? Do they motivate amount of work or actual material learned?
9. Do marks motivate parents as well, or what is the relationship between home and school in regard to marks?
10. Consider the following situation: A year end marking in which more than three quarters of the 30 students in Teacher X's class received C or D? What would you conclude? What additional information would you need to come to a conclusion?

Interview #3 - June

1. May I please record your final marks.
2. Would you please do two quick sorts for me.
 - A. Please sort the class into two categories according to effort: those who put forth average to above average on a regular basis and those who put forth average and below.
 - B. Please sort the class into two categories according to ability: those whom you perceive to be generally average to above average ability, those whom you perceive to be generally average to below average ability.
3. Your class averages above a B in language arts and in math. Some people think that classes should average a C. Could you discuss some of your reasons why your class is higher.
4. You have no (one) E's so nobody has failed. Do you have a theory about marks below C?
5. If you had to make a quick judgment about your own marking, which would you say carries the most weight in deciding a mark - the effort or the outcome? why?
6. In several studies of classroom achievement, four factors were repeatedly mentioned by teacher as being regularly influential - ability, effort, task difficulty, luck. You have frequently mentioned ability and effort, but you have added several distractions such as lack of ability to concentrate, carelessness, low level of home support or divorce, and interest in socializing. How do these affect marks?
7. Would you mind a follow up question during the summer? Phone number.

APPENDIX B

REPORT CARD 1

REPORT CARD 2

APPENDIX B

Report Card Which was Used in the School of Teachers One and Two

ELEMENTARY SCHOOL

EVALUATION OF
STUDENT PROGRESS

	1	2	3	4	FINAL
MATHEMATICS					
Grade level					
Completes assignments					

LANGUAGE ARTS					
Communicates well orally					
Expresses self in written form					
Uses correct grammar					

READING LEVEL					
Grade level					
Reads with expression					
Completes assignments					
Comprehends materials					
Uses reading skills					

EXPLANATION OF GRADES

- A - Excellent achievement
 B - Above average achievement
 C - Average achievement
 D - Below average achievement
 E - Unacceptable achievement
 I - Incomplete work
 No mark - Not applicable at this time
 * - Refer to comments

	1	2	3	4	FINAL
SCIENCE					

SOCIAL STUDIES					
-----------------------	--	--	--	--	--

SPELLING					
Grade level					
Spells assigned words correctly					
Uses correct spelling in other subject areas					
Completes assignments					

KEY TO SUBHEADINGS

- ✓ - Needs Improvement
 No ✓ - Satisfactory Work

	1	2	3	4	FINAL
PENMANSHIP					
Neat and legible penmanship					
Neat and legible penmanship in other subject areas					

CITIZENSHIP					
Is courteous					
Is considerate of the rights of others					
Controls talking					

STUDY HABITS					
Works independently					
Listens carefully and follows directions					
Uses time wisely					
Works neatly					
Completes assignments on time					

Report Card Which was Used in the School of Teachers Three, Four and Five

Oakland County, Michigan

Pupil _____

Grade _____ School Year _____

Teacher _____

School _____

DEVELOPMENT KEY

- 1 - Outstanding Achievement
2 - Satisfactory Achievement
3 - More Growth Or Achievement is Necessary
0 - Indicates The Area is Not Relevant To This Marking Period

KEY TO GRADES:

- REF TO
A - EXCELLENT
B - ABOVE AVERAGE
C - AVERAGE
D - BELOW AVERAGE
E - FAILING

SUBJECT	MARKING PERIOD 1 2 3 4 F
LANGUAGE	
READING	
Achieve Oral from Expression	
Understands what is Read	
Journals how things independently	
Uses Reference and Study Skills	
Completes Assignments	
Makes Independent for Progress	
LANGUAGE	
Know Correct Grammar	
Expresses ideas Clearly in Writing	
Expresses Ideas Clearly in Speaking	
Use Dictionary	
Completed Assignments	
SPELLING	
Know Assigned Words	
Write Other Words from Context	
Use Dictionary	
Completed Assignments	
HANDWRITING	
Forms Letters Correctly	
Writes Legibly	
Completed Assignments	

SUBJECT	MARKING PERIOD 1 2 3 4 F
MATHEMATICS	
Understands Place Value	
Knows Basic Number Facts	
Recognizes What is Story Problems	
Is Accurate in Steps	
Understands Concepts Being Covered	
Completes Assignments	
SOCIAL STUDIES	
Understands Basic Concepts	
Uses Necessary Map and Globe Skills	
Contributes to Activities & Discussion	
Tells Performance	
Shows Interest and Enthusiasm	
Uses Reference Materials	
Completes Assignments	
SCIENCE	
Understands Basic Concepts	
Observes, Measures & Records Accurately	
Shows Interest and Enthusiasm	
Completes Assignments	

SUBJECT	MARKING PERIOD 1 2 3 4 F
SOCIAL BEHAVIOR	
is Courteous	
Respects Property of Others	
Respects Authority	
Follows Classroom Rules	
Follows School Rules	
Appears Neat and Clean	
Carefully Work Related and Alert	
Exhibits Self Discipline	
WORK HABITS	
is Arranged	
Follows Directions	
Values Good Use of Class Time	
Works Hard When Needed	
Values Work in Group	
Views their independence	
Always Neatly	
is Well Organized	
ATTENDANCE	
misses Tardy	
Late Absent	

COMMENTS

1

2

3

4

5

APPENDIX C

Attribution-Utility Coding Device

This instrument was the result of the categorizing of teacher comments taken from the transcribed interviews.

Home support level

	1	2	3	4	5	All
Supportive						
Parents very responsive to need for work.						
Parents very responsive to need for skill.						
Father especially responsive.						
Mother especially responsive.						
Parent's absolutely stated that it wasn't all B's.						
"I really love the parents up."						
Aunt and Uncle who really care.						
Parents will be sure the marks are A's.						
Total supportive						
Problematic (often leading to poor study habits).						
Ill parents.						
Death of a parent.						
Recent divorce, <i>split family</i>						
Recent re-marriage.						
Language problems (second language).						
Single parent raising home.						
Both parents working, too tired for discipline.						
Elderly parents without much energy.						
Father left the home, anger.						
Mother has had several husbands, name change.						
Slacer on drugs, hospitalized.						
Total problematic						
Unsupportive - <i>Support lacking</i>						
Mother ran his down so badly.						
Mother says he is mentally retarded, he isn't.						
Absence or tardiness excessive without illness or excuse.						
Punitive, ridiculous penalties						
Total home support.						
	17	2	14	2	9	51

Classroom Behavior/Security/Developmental

	1	2	3	4	5	All
Physical						
Growth Spurt, growing rapidly.						
Very large, heavy, big for age.						
Small for age.						
Hard time with himself.						
Puberty.						
On medication.						
Can't sit still long enough to do anything.						
Total Physical.						
Social						
Very withdrawn.						
New Socialite.						
Interested in girls, hair, etc.						
Lady's man/boy crazy.						
Reluctant, likes to visit.						
Flighty, can't settle.						
Total social.						
Emotional						
Emotional problems, personal problems.						
Very, very sensitive.						
Constantly worried.						
Very immature.						
Very mature and dependable.						
Always helps underdog, kind.						
Likes to please others.						
Likes to please me (teacher).						
Tells out answers, lacks control.						
Nervous problems.						
Total emotional.						
Total Behavior.	7	15	6	12	9	49

Although each of these comments is quoted directly from the teachers, for coding attribution, only one comment per category was allowed. When several comments were used to describe one student's success or failure, those above one were included in the explanation about elaboration. Effort, home support and classroom behavior were much elaborated categories.

TEACHER ATTRIBUTION COMMENTS		Teacher One	Teacher Two	Teacher Three	Teacher Four	Teacher Five	Composite
Ability (achievement)							
Upper Elementary Students 152							
Concept of average: above, good, much below, low, below grade level.							
Concept of bright: very bright, abnormally top notch student, brightest kid in class, slow, dull.							
Concept of achiever: over/under, high/low.							
Concept of grades: A, B, C student, D-C, straight A student.							
Fast Freese.							
Special Education student.							
Total ability.		27	21	21	21	27	117
Effort (motivation)							
In class:							
Attending, concentration.							
Wasting time, laziness, lack of discipline.							
Speed, carelessness, finishes in 5 minutes.							
Total in class							
Out of class:							
Conscientious, works hard.							
Has poor study/work habits.							
Does extra work, more than is asked for.							
Makes up all assignments.							
Works ahead.							
Total out of class							
Comprehensive:							
Over achieving, really trying.							
Underachieving, unstable effort.							
Competitive, keeps up with friends.							
"Stimulating him to do anything is almost a case to case basis".							
Determined to get all A's.							
Can't get his act together.							
No motivation.							
Not much enthusiasm.							
Needs to be prodded constantly.							
Very disorganized.							
Total effort.		26	9	24	10	21	90
Task difficulty							
Skills							
Multiplication not mastered							
Division is often difficult.							
Problem expressing ideas in writing.							
Students speak well so we're working on writing.							
Addition and subtraction not mastered.							
May die as concepts become more difficult.							
Total							
Text Book level							
Reading above grade level.							
Reading at grade level.							
Reading a couple of grades below level.							
Social Studies book is difficult							
Social Studies tests are hard.							
Total							
General							
Learning disabled.							
Child is being tested.							
Has a hard time learning my goals.							
Trouble concentrating.							
Better in language.							
Better in math.							
Discusses well.							
Total - Total task difficulty.			10		9	19	

APPENDIX D

CASE STUDY CHARTS

APPENDIX D

Case Study Charts

The data contained in these tables were used to modify the original regression equation. Similar tables for the composite case were included in the body of the text, with interpretation, as explanation of the adjustment process.

Teachers One, Two,
Three, Four and Five }

Original Regression Equation
Regression Tables, Pearson
Correlations, Partial

APPENDIX D

CASE STUDY CHARTS

TEACHER ONE

Variable Weights Within the Final Mark

Teacher One

Dependent Variable = L ₅ Final Language Mark				
Language Arts Variables		B	F	Sign.
Final Prediction	L ₄	.60	5.35	.030
Second Marking	L ₃	.34	3.5	.072
Second Prediction	L ₂	.066	.057	.812
First Marking	L ₁	.045	.029	.865
Note: Overall F = 41.97 Multiple R = .94 R Square = .87 Standard Deviation = 1.42				

Dependent Variable = M ₅ Final Math Mark				
Math Variables		B	F	Sign.
Final Prediction	M ₄	.096	.073	.788
Second Marking	M ₃	.69	5.28	.030
Second Prediction	M ₂	.11	.145	.707
First Marking	M ₁	.16	.436	.515
Note: Overall F = 43. Multiple R = .94 R Square = .87 Standard Deviation = 1.40				

Note. These weights are derived through regression analysis of the numerical equivalents of grades of an upper elementary class across a year.

Correlations Among Markings and Predictions

Teacher One

Language Arts Variables		L ₁	L ₂	L ₃	L ₄	L ₅
First Marking	L ₁	1,000	.95	.73	.77	.75
Second Prediction	L ₂		1,000	.70	.76	.74
Second Marking	L ₃			1,000	.93	.91
Final Prediction	L ₄				1,000	.92
Final Marking	L ₅					1,000

P=.001

Mathematics Variables		M ₁	M ₂	M ₃	M ₄	M ₅
First Marking	M ₁	1,000	.94	.85	.85	.85
Second Prediction	M ₂		1,000	.84	.88	.84
Second Marking	M ₃			1,000	.96	.93
Final Prediction	M ₄				1,000	.91
Final Marking	M ₅					1,000

P=.001

Regression Equations:

$$L_5 = .60L_4 + .35L_3 - \text{Constant}$$

(Sign. = .030) (Sign. = .072)

$$M_5 = .80M_3 + .28M_2 - \text{Constant}$$

(Sign. = .000) (Sign. = .083)

Note. These correlations are based on the language arts and mathematics marks and predictions of an upper elementary class across one year.

Partial Correlation Coefficients

Teacher One

Language Arts	Controlling for:	Coefficient	P
$r(L_4 L_5)$	L_1	.82	.001
	L_2	.83	.001
	L_3	.51	.003
	$L_1 L_2$.82	.001
	$L_1 L_3$.44	.011
	$L_2 L_3$.43	.013
	$L_1 L_2 L_3$.43	.015

$r(L_2 L_3)$	L_1	.03	.439
--------------	-------	-----	------

Note. L_1 = First Actual Mark
 L_2 = Second Prediction
 L_3 = Second Actual Mark
 L_4 = Final Prediction
 L_5 = Final Mark

Mathematics	Controlling for:	Coefficient	P
$r(M_4 M_5)$	M_1	.67	.001
	M_2	.65	.001
	M_3	.18	.179
	$M_1 M_2$.64	.001
	$M_1 M_3$.11	.286
	$M_2 M_3$.01	.471
	$M_1 M_2 M_3$.05	.394

$r(M_2 M_3)$	M_1	.21	.136
--------------	-------	-----	------

Note. M_1 = First Actual Mark
 M_2 = Second Prediction
 M_3 = Second Actual Mark
 M_4 = Final Prediction
 M_5 = Final Mark

Note. Partial correlation coefficients were derived from the numerical equivalents of the marks of a class of upper elementary students across a year.

APPENDIX D

CASE STUDY CHARTS

TEACHER TWO

Variable Weights Within the Final Mark

Teacher Two

Dependent Variable = L ₅ Final Language Mark				
Language Arts Variables		B	F	Sign.
Final Prediction	L ₄	.32	2.08	.162
Second Marking	L ₃	.34	2.74	.111
Second Prediction	L ₂	.040	.120	.732
First Marking	L ₁	.13	1.28	.268
Note: Overall F = 12.94 Multiple R = .83 R Square = .69 Standard Deviation = .90				

Dependent Variable = M ₅ Final Math Mark				
Math Variables		B	F	Sign.
Final Prediction	M ₄	.40	4.91	.037
Second Marking	M ₃	.009	.011	.916
Second Prediction	M ₂	.052	.026	.613
First Marking	M ₁	.18	3.88	.061
Note: Overall F = 18. Multiple R = .87 R Square = .75 Standard Deviation = .82				

Note. These weights are derived through regression analysis of the numerical equivalents of grades of an upper elementary class across a year.

Correlations Among Markings and Predictions

Teacher Two

Language Arts Variables		L ₁	L ₂	L ₃	L ₄	L ₅
First Marking	L ₁		.62	.67	.65	.68
Second Prediction	L ₂			.49 P=.004	.57	.53 P=.002
Second Marking	L ₃				.83	.78
Final Prediction	L ₄					.78
Final Marking	L ₅					

P=.001

Mathematics Variables		M ₁	M ₂	M ₃	M ₄	M ₅
First Marking	M ₁		.81	.51 P=.003	.75	.81
Second Prediction	M ₂			.48 P=.005	.71	.74
Second Marking	M ₃				.74	.60
Final Prediction	M ₄					.82
Final Marking	M ₅					

P=.001

Regression Equations:

$$L_5 = .40L_2 + .34L_3 + .32L_4 - \text{Constant}$$

(Sign. = .732) (Sign. = .111) (Sign. = .162)

$$M_5 = .42M_4 + .20M_1 - \text{Constant}$$

(Sign. = .003) (Sign. = .007)

Note. These correlations are based on the language arts and mathematics marks and predictions of an upper elementary class across one year.

Partial Correlation Coefficients

Teacher Two

Language Arts	Controlling for:	Coefficient	P
$r(L_4 L_5)$	L_1	.60	.001
	L_2	.68	.001
	L_3	.37	.029
	$L_1 L_2$.58	.001
	$L_1 L_3$.32	.058
	$L_2 L_3$.30	.067
	$L_1 L_2 L_3$.29	.081

$r(L_2 L_3)$	L_1	.12	.274
--------------	-------	-----	------

Note. L_1 = First Actual Mark
 L_2 = Second Prediction
 L_3 = Second Actual Mark
 L_4 = Final Prediction
 L_5 = Final Mark

Mathematics	Controlling for:	Coefficient	P
$r(M_4 M_5)$	M_1	.55	.001
	M_2	.62	.001
	M_3	.70	.001
	$M_1 M_2$.52	.003
	$M_1 M_3$.45	.011
	$M_2 M_3$.51	.004
	$M_1 M_2 M_3$.42	.018

$r(M_2 M_3)$	M_1	.55	.001
--------------	-------	-----	------

Note. M_1 = First Actual Mark
 M_2 = Second Prediction
 M_3 = Second Actual Mark
 M_4 = Final Prediction
 M_5 = Final Mark

Note. Partial correlation coefficients were derived from the numerical equivalents of the marks of a class of upper elementary students across a year.

APPENDIX D

CASE STUDY CHARTS

TEACHER THREE

Variable Weights Within the Final Mark

Teacher Three

Dependent Variable = L ₅ Final Language Mark				
Language Arts Variables		B	F	Sign.
Final Prediction	L ₄	.031	.011	.916
Second Marking	L ₃	.38	2.75	.111
Second Prediction	L ₂	.014	.03	.957
First Marking	L ₁	.69	6.72	.017
Note: Overall F = 53. Multiple R = .95 R Square = .90 Standard Deviation = 1.22				

Dependent Variable = M ₅ Final Math Mark				
Math Variables		B	F	Sign.
Final Prediction	M ₄	.36	1.9	.173
Second Marking	M ₃	.14	.35	.556
Second Prediction	M ₂	-.16	.52	.477
First Marking	M ₁	.46	10.9	.003
Note: Overall F = 38. Multiple R = .93 R Square = .87 Standard Deviation = 1.16				

Note. These weights are derived through regression analysis of the numerical equivalents of grades of an upper elementary class across a year.

Correlations Among Markings and Predictions
Teacher Three

Language Arts Variables		L ₁	L ₂	L ₃	L ₄	L ₅
First Marking	L ₁		.94	.82	.89	.92
Second Prediction	L ₂			.83	.87	.89
Second Marking	L ₃				.93	.89
Final Prediction	L ₄					.91
Final Marking	L ₅					

P=.001

Mathematics Variables		M ₁	M ₂	M ₃	M ₄	M ₅
First Marking	M ₁		.89	.88	.84	.91
Second Prediction	M ₂			.86	.90	.84
Second Marking	M ₃				.93	.89
Final Prediction	M ₄					.87
Final Marking	M ₅					

P=.001

Regression Equations:

$$L_5 = .69L_1 + .38L_3 - \text{Constant}$$

(Sign. = .017) (Sign. = .111)

$$M_5 = .45M_1 + .38M_4 - \text{Constant}$$

(Sign. = .000) (Sign. = .011)

Note. These correlations are based on the language arts and mathematics marks and predictions of an upper elementary class across one year.

Partial Correlation Coefficients

Teacher Three

Language Arts	Controlling for:	Coefficient	P
$r(L_4 L_5)$	L_1	.50	.003
	L_2	.61	.001
	L_3	.49	.004
	$L_1 L_2$.49	.004
	$L_1 L_3$.06	.388
	$L_2 L_3$.29	.064
	$L_1 L_2 L_3$.05	.394

$r(L_2 L_3)$	L_1	.28	.065
--------------	-------	-----	------

Note: L_1 = First Actual Mark
 L_2 = Second Prediction
 L_3 = Second Actual Mark
 L_4 = Final Prediction
 L_5 = Final Mark

Mathematics	Controlling for:	Coefficient	P
$r(M_4 M_5)$	M_1	.49	.006
	M_2	.50	.005
	M_3	.27	.091
	$M_1 M_2$.48	.007
	$M_1 M_3$.25	.117
	$M_2 M_3$.13	.268
	$M_1 M_2 M_3$.29	.086

$r(M_2 M_3)$	M_1	.40	.022
--------------	-------	-----	------

Note: M_1 = First Actual Mark
 M_2 = Second Prediction
 M_3 = Second Actual Mark
 M_4 = Final Prediction
 M_5 = Final Mark

Note. Partial correlation coefficients were derived from the numerical equivalents of the marks of a class of upper elementary students across a year.

APPENDIX D

CASE STUDY CHARTS

TEACHER FOUR

Variable Weights Within the Final Mark

Teacher Four

Dependent Variable = L ₅ Final Language Mark				
Language Arts Variables		B	F	Sign.
Final Prediction	L ₄	-.55	1.68	.206
Second Marking	L ₃	.93	6.19	.019
Second Prediction	L ₂	.43	3.78	.062
First Marking	L ₁	.030	.035	.854
Note: Overall F = 36. Multiple R = .91 R Square = .84 Standard Deviation = 1.04				

Dependent Variable = M ₅ Final Math Mark				
Math Variables		B	F	Sign.
Final Prediction	M ₄	.15	.63	.434
Second Marking	M ₃	.38	3.08	.091
Second Prediction	M ₂	.21	1.5	.227
First Marking	M ₁	.18	1.4	.243
Note: Overall F = 57. Multiple R = .95 R Square = .89 Standard Deviation = 1.05				

Note. These weights are derived through regression analysis of the numerical equivalents of grades of an upper elementary class across a year.

Correlations Among Markings and Predictions

Teacher Four

Language Arts Variables		L ₁	L ₂	L ₃	L ₄	L ₅
First Marking	L ₁		.88	.85	.84	.82
Second Prediction	L ₂			.91	.93	.88
Second Marking	L ₃				.98	.90
Final Prediction	L ₄					.88
Final Marking	L ₅					

P=.001

Mathematics Variables		M ₁	M ₂	M ₃	M ₄	M ₅
First Marking	M ₁		.93	.87	.83	.89
Second Prediction	M ₂			.89	.83	.90
Second Marking	M ₃				.94	.93
Final Prediction	M ₄					.89
Final Marking	M ₅					

P=.001

Regression Equations:

$$L_5 = .93L_3 + .43L_2 - \text{Constant}$$

(Sign. = .019) (Sign. = .062)

$$M_5 = .59M_3 + .31M_1 - \text{Constant}$$

(sign. = .000) (Sign. = .009)

Note. These correlations are based on the language arts and mathematics marks and predictions of an upper elementary class across one year.

Partial Correlation Coefficients

Teacher Four

Language Arts	Controlling for:	Coefficient	P
$r(L_4 L_5)$	L_1	.60	.001
	L_2	.33	.033
	L_3	-.03	.431
	$L_1 L_2$.32	.042
	$L_1 L_3$	-.05	.388
	$L_2 L_3$	-.25	.086
	$L_1 L_2 L_3$	-.242	.103

$r(L_2 L_3)$	L_1	.64	.001
--------------	-------	-----	------

Note. L_1 = First Actual Mark
 L_2 = Second Prediction
 L_3 = Second Actual Mark
 L_4 = Final Prediction
 L_5 = Final Mark

Mathematics	Controlling for:	Coefficient	P
$r(M_4 M_5)$	M_1	.58	.001
	M_2	.59	.001
	M_3	.13	.240
	$M_1 M_2$.54	.001
	$M_1 M_3$.12	.270
	$M_2 M_3$.18	.172
	$M_1 M_2 M_3$.15	.217

$r(M_2 M_3)$	M_1	.43	.008
--------------	-------	-----	------

Note. M_1 = First Actual Mark
 M_2 = Second Prediction
 M_3 = Second Actual Mark
 M_4 = Final Prediction
 M_5 = Final Mark

Note. Partial correlation coefficients were derived from the numerical equivalents of the marks of a class of upper elementary students across a year.

APPENDIX D

CASE STUDY CHARTS

TEACHER FIVE

Variable Weights Within the Final Mark

Teacher Five

Dependent Variable = L ₅ Final Language Mark				
Language Arts Variables		B	F	Sign.
Final Prediction	L ₄	.52	3.64	.067
Second Marking	L ₃	.090	.119	.732
Second Prediction	L ₂	.42	8.1	.008
First Marking	L ₁	-.08	.016	.684
Note: Overall F = 43. Multiple R = .93 R Square = .87 Standard Deviation = 1.06				

Dependent Variable = M ₅ Final Math Mark				
Math Variables		B	F	Sign.
Final Prediction	M ₄	.56	7.64	.010
Second Marking	M ₃	.45	4.3	.048
Second Prediction	M ₂	-.122	.49	.488
First Marking	M ₁	.048	.065	.799
Note: Overall F = 29. Multiple R = .90 R Square = .81 Standard Deviation = 1.20				

Note. These weights are derived through regression analysis of the numerical equivalents of grades of an upper elementary class across a year.

Correlations Among Markings and Predictions

Teacher Five

Language Arts Variables		L ₁	L ₂	L ₃	L ₄	L ₅
First Marking	L ₁		.90	.92	.92	.87
Second Prediction	L ₂			.85	.87	.90
Second Marking	L ₃				.95	.88
Final Prediction	L ₄					.91
Final Marking	L ₅					

P=.001

Mathematics Variables		M ₁	M ₂	M ₃	M ₄	M ₅
First Marking	M ₁		.78	.76	.69	.66
Second Prediction	M ₂			.74	.71	.63
Second Marking	M ₃				.90	.87
Final Prediction	M ₄					.88
Final Marking	M ₅					

P=.001

Regression Equations:

$$L_5 = .90L_3 + .52L_4 + .42L_2 - \text{Constant}$$

(Sign. = .732) (Sign. = .067) (Sign. = .008)

$$M_5 = .53M_4 + .42M_3 - \text{Constant}$$

(Sign. = .010) (Sign. = .035)

Note. These correlations are based on the language arts and mathematics marks and predictions of an upper elementary class across one year.

Partial Correlation Coefficients

Teacher Five

Language Arts	Controlling for:	Coefficient	P
$r(L_4 L_5)$	L_1	.56	.001
	L_2	.58	.001
	L_3	.48	.003
	$L_1 L_2$.51	.002
	$L_1 L_3$.42	.012
	$L_2 L_3$.34	.034
	$L_1 L_2 L_3$.35	.034

$r(L_2 L_3)$	L_1	.13	.235
--------------	-------	-----	------

Note. L_1 = First Actual Mark
 L_2 = Second Prediction
 L_3 = Second Actual Mark
 L_4 = Final Prediction
 L_5 = Final Mark

Mathematics	Controlling for:	Coefficient	P
$r(M_4 M_5)$	M_1	.79	.001
	M_2	.80	.001
	M_3	.46	.005
	$M_1 M_2$.77	.001
	$M_1 M_3$.46	.006
	$M_2 M_3$.48	.005
	$M_1 M_2 M_3$.47	.005

$r(M_2 M_3)$	M_1	.79	.001
--------------	-------	-----	------

Note. M_1 = First Actual Mark
 M_2 = Second Prediction
 M_3 = Second Actual Mark
 M_4 = Final Prediction
 M_5 = Final Mark

Note. Partial correlation coefficients were derived from the numerical equivalents of the marks of a class of upper elementary students across a year.

APPENDIX E

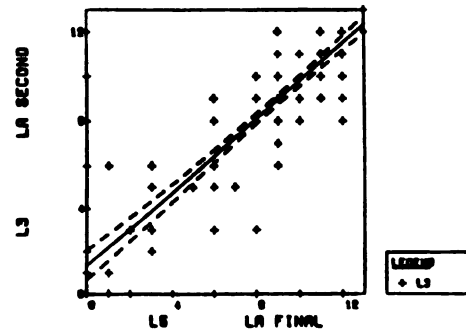
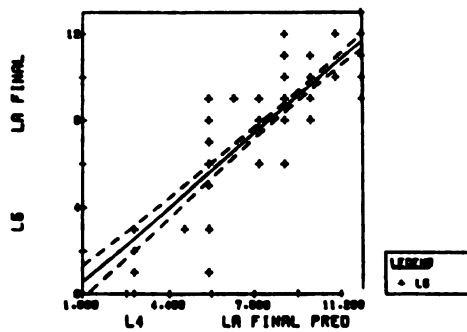
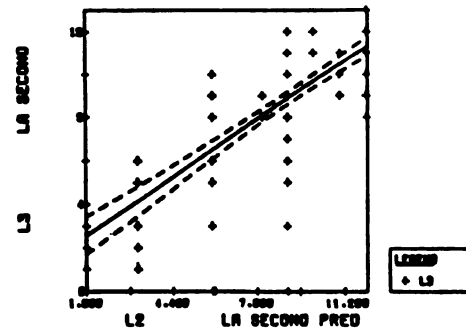
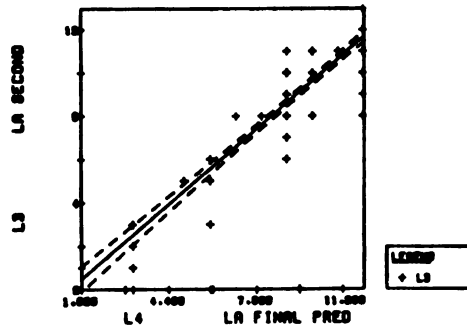
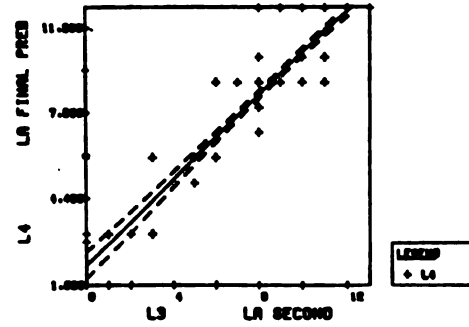
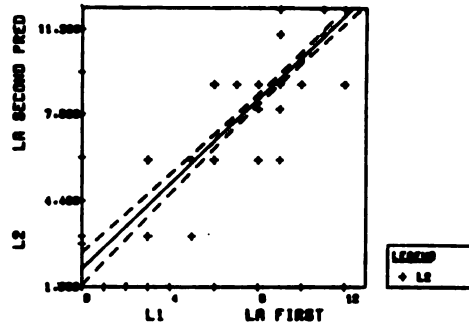
COMPOSITE REGRESSION PLOTS

LANGUAGE

MATHEMATICS

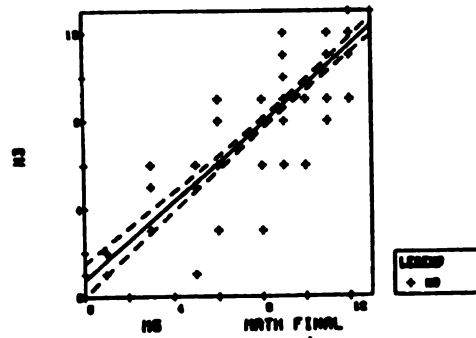
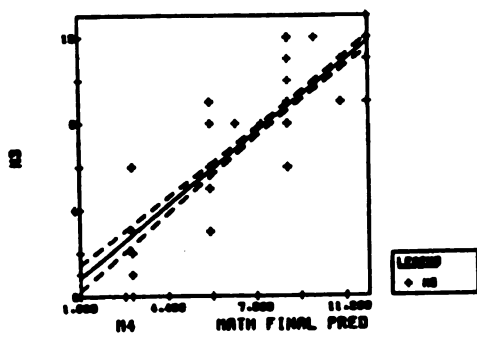
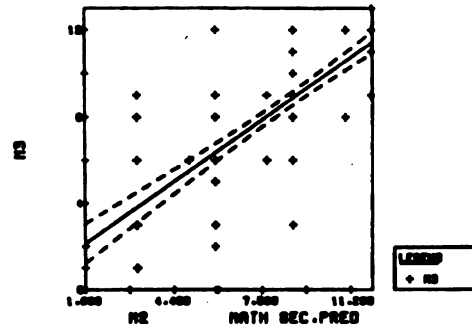
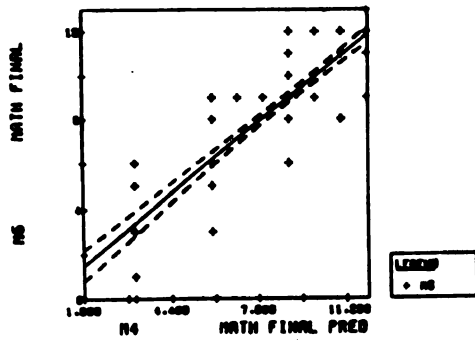
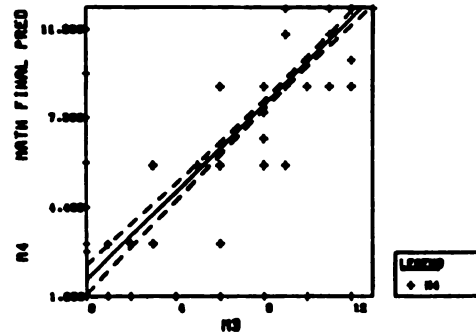
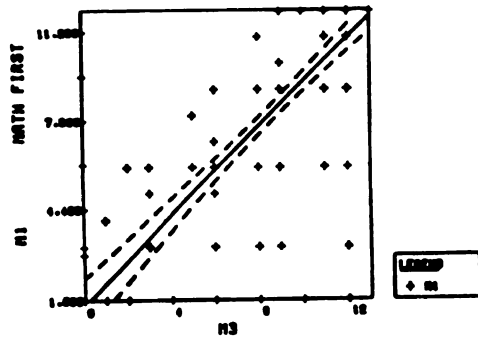
APPENDIX E

COMPOSITE REGRESSION PLOTS - LANGUAGE



APPENDIX E

COMPOSITE REGRESSION PLOTS - MATHEMATICS



APPENDIX F

COMPOSITE CROSS TABULATION

APPENDIX F

Composite Cross Tabulations

All Teachers

		EFFORT		
		High	Low	
ABILITY	High	93	24	117
	Low	9	24	33
		102	48	150

LIST OF REFERENCES

LIST OF REFERENCES

- Aiken, L. The grading behavior of a college faculty. Educational and Psychological Measurement, 1963, 23, 319-322.
- Anderson, R. The importance and purpose of reporting. The National Elementary Principal, May 1966, XLV, No. 6.
- Baird, L., & Feister, W. Grading standards: The relation of changes in average student ability to the average grades awarded. American Educational Research Journal, Summer 1972, 9, 440.
- Bandura, A. Social Learning Theory. New Jersey: Prentice-Hall, 1977.
- Barnes, R., Ickes, W., & Kidd, R. Effects of perceived intentionality and stability of another's dependency on helping behavior: A field experiment. Unpublished manuscript, University of Wisconsin, Madison, 1977. Cited by B. Weiner, 1979.
- Becker, H., Geer, B., & Hughes, E. Making the Grade: The Academic Side of College Life. New York: Wiley, 1968.
- Bejar, I., & Blew, E. Grade inflation and the validity of the Scholastic Aptitude Test. American Educational Research Journal, Summer 1981, 18, No. 2, 143-156.
- Bells, W. Reliability of repeated grading of essay type examination. Journal of Educational Psychology, 1930, 21, 48-52.
- Berkowitz, L. Resistance to improper dependency relationships. Journal of Experimental Social Psychology, 1969, 5, 283-294.
- Bloom, B. Human Characteristics and School Learning. New York: McGraw-Hill, 1976.
- Bloom, B. Taxonomy of Educational Objectives, Handbook I: Cognitive Domain. New York: David McKay Company, Inc., 1956.
- Brophy, J. Teachers' cognitive activities and overt behaviors. Occasional Paper, Institute for Research on Teaching, Michigan State University, April 1980.
- Brophy, J. Recent research on teaching. Occasional Paper No. 40, Institute for Research on Teaching, Michigan State University, November 1980.
- Brophy, J., & Good, T. Teachers communication of differential expectations for children's classroom performance: Some behavioral data. Journal of Educational Psychology, 1970, 61, 365-379.
- Brophy, J., & Good, T. Teacher-Student Relationships. New York: Holt, Rinehart and Winston, Inc., 1974.

- Broudy, H. The Real World of the Public Schools. New York: Harcourt Brace Jovanovich, Inc., 1972.
- Bussis, A., Chittenden, E., & Amarel, M. Beyond Surface Curriculum. Colorado: Westview Press, 1976.
- Carroll, J., & Payne, J. (Ed.) Cognition and Social Behavior. New Jersey: Erlbaum Associates, 1976. (Distributed by John Wikz and Sons, New York.)
- Carroll, J., & Payne, J. The psychology of the parole decision process: A joint application of attribution theory and information processing psychology. Cognition and Social Behavior. New Jersey: Erlbaum Associates, 1976.
- Chansky, N. The x-ray of the school mark. Educational Forum, March 1962, 347-352.
- Clark, C., & Yinger, R. Research on teacher thinking. Institute for Research on Teaching, Michigan State University, Research Series No. 12, April 1978.
- Clark, C., Yinger, R., & Wildfong, S. Identifying cues for use in studies of teacher judgment. Institute for Research on Teaching, Michigan State University, Research Series No. 23, July 1978.
- Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfield, F., & York, R. Equality of Educational Opportunity. U.S. Office of Health, Education and Welfare, 1966.
- Coleman, J., Hoffer, T., & Kilgore, S. Public and private schools: A report to the National Center for Education. National Opinion Research Center, University of Chicago, March 1981.
- Combs, A. Educational Accountability Beyond Behavioral Objectives. Washington: Association for Supervision and Curriculum Development, 1972.
- Cooper, H., & Burger, J. Internality, stability and personal efficacy: A categorization of free response academic attributions. Unpublished manuscript, University of Missouri, Columbia, 1978.
- Craik, F. Primary Memory. British Medical Bulletin, 1971. 27, 232-236.
- Cremin, L. The Genius of American Education. New York: Vintage Books, 1965.
- Cremin, L. The Transformation of the School. New York: Vintage Books, 1961.
- Cronbach, L. Beyond the two disciplines of scientific psychology. American Psychologist, 1957, 12, 671-684.
- Dearborn, W. School and University Grades. University of Wisconsin, 1910.
- Doyle, W. Paradigms for research on teacher effectiveness. In L. Shulman (Ed.). Review of Research in Education, 5, Illinois: Peacock, 1977.

- Doyle, W. Student mediating responses in teaching effectiveness: Final Report. (Project No. 0-0645.) Department of Education, North Texas University, March 1980.
- Dunkin, M., & Biddle, B. The Study of Teaching. New York: Holt, Rinehart and Winston, Inc., 1974.
- Dusek, J. Do teachers bias children's learning? Review of Educational Research, 1975, 45, 661-684.
- Educational Research Service. Reporting Pupil Progress: Policies, Procedures and Systems. Virginia: 1977.
- Einhorn, H., & Hogarth, R. Confidence in judgment: Persistence of illusion of validity. Psychological Review, 1978, 85, No. 5.
- Einhorn, H., Kleinmuntz, B., & Kleinmuntz, D. Linear Regression and Process Tracing Models of Clinical Judgment. Unpublished paper, Graduate School of Business, University of Chicago, January 1979.
- Ericsson, K., & Simon, H. Verbal reports as data. Psychological Review, May 1980, 87, No. 3.
- Evans, F. What research says about grading. S. Simon & J. Bellanca, (Eds.). Washington: Association of Supervision and Curriculum Development, 1976.
- Fay, P. The effect of the knowledge of marks on the subsequent achievement of college students. Journal of Educational Psychology, 1937, 28, 548-554.
- Ferguson, R., & Maxey, E. Trends in academic performance of high school and college students. Paper, 1975. Available from Educational Resources and Information Center, U.S. Office of Education, ED 109 523.
- Fischhoff, B. Attribution theory and judgment under uncertainty. New Directions in Attribution Research. New Jersey: Erlbaum Associates, 1976.
- Fischhoff, B., & Beyth, R. I knew it would happen. Remembered probabilities of once—future things. Organizational Behavior and Human Performance, 1975, 13.
- Freyd, M. The graphic scale. Journal of Educational Psychology, 1923, 14, 94.
- Gage, N. The Scientific Basis of the Art of Teaching. New York: Teachers College Press, Columbia University, 1978.
- Getzels, J., & Jackson, P. The teacher's personality and characteristics. In N. L. Gage (Ed.). Handbook of Research on Teaching. Chicago: Rand McNally, 1963.
- Goldberg, L. Man versus model of man: A rationale, plus some evidence, for a method of improving clinical inferences. Psychological Bulletin, 1970, 73, No. 6.

- Heider, F. The Psychology of Interpersonal Relations. New York: Wiley, 1958.
- Hilsom, S., & Cane, B. The Teacher's Day. London: National Foundation for Educational Research in England and Wales, 1971.
- Jaggard, G. Improving the marking system. Educational Administrative Supplement, 1919, 5, 25-35.
- Johnson, D. The Psychology of Thought and Judgment. New York: Harper Row, 1955.
- Joyce, B. Toward a theory of information processing in teaching. Educational Research Quarterly, 1978-79, 3, 73-77.
- Kahneman, D., & Tversky, A. On the psychology of prediction. Psychological Review, July 1973, 80, No. 4.
- Kahneman, D., & Tversky, A. Subjective probability. Cognitive Psychology, July 1972, 3, No. 3.
- Kelly, F. Teachers' marks, their variability and standardization. Teacher's College Record, Columbia University, New York: 1914.
- Kelley, H. Attribution Theory in Social Psychology. New Jersey: General Learning Press, 1971. (Based on a previous paper of the same title.) In D. Levine (Ed.) Nebraska Symposium on Motivation. Lincoln: University of Nebraska Press, 1967.
- Kirschenbaum, H., Simon, S., & Napier, R. Wad-Ja-Get? The Grading Game in American Education. New York: Hart Publishing Co., 1971.
- Lavin, D. The Prediction of Academic Performance. New York: The Russell Sage Foundation, 1965.
- Lessinger, L. After Texarkana, what? Nations Schools, December 1969, 84, Vol. 6, 37-40.
- Lezotte, L., Miller, D., Passalacqua, J., & Brookover, W. School learning climate and student achievement. Document: SSTA Center, c/o Teacher Education Project, Florida State University, 403 Education Building, Tallahassee, Florida 32306, 1976.
- Long, R., & Henderson, E. The effect of pupils' race, class, test scores, and classroom behavior on the academic expectancies of southern and nonsouthern white teachers. Paper presented at annual meeting of American Educational Research Association, 1972. Cited by J. Brophy, Teachers cognitive activities and overt behavior.
- Los Angeles Committee of the Secondary School Principals Association. Marking slow pupils. California Quarterly of Secondary Education, 1926, 1, 386-391. Cited in Smith and Dobbins, Encyclopedia of Educational Research, 1957.

- Mackay, D., & Marland, P. Thought Processes of Teachers. Paper presented at the annual meeting of the American Educational Research Association. Toronto: February 1978.
- Marland, P. A Study of Teachers' Interactive Thoughts. Unpublished doctoral dissertation, University of Alberta, 1977.
- Marx, R. Teacher Judgments of Students' Cognitive and Affective Outcomes. Unpublished doctoral dissertation, Stanford University, 1978.
- McDavid, J. Some relationships between social reinforcement and scholastic achievement. Journal of Consulting Psychology, 1959, 23, 151-154.
- Medley, D. The effectiveness of teachers. In P. Peterson and H. Walberg (Eds.). Research on Teaching: Concepts, Findings and Implications. California: McCutchan, 1979.
- Meichenbaum, D. Cognitive Behavior Modification. New York: Plenum, 1977. Cited in J. Brophy, Teacher Cognitive Activities and Overt Behaviors.
- Miller, G. The magical number seven plus or minus two: Some limits on our capacity for processing information. Psychological Review, 1956, 63, 81-97.
- Morine-Dersheimer, G. Teachers' conceptions of pupils--An outgrowth of instructional context. The South Bay Study, Part III. Institute for Research on Teaching, Michigan State University, Research Series No. 59, July 1978.
- National Institute of Education. Teaching as Clinical Information Processing. Washington: National Conference on Studies in Teaching (Report of Panel 6), 1975.
- National Education Association. What Research Says to the Teacher: Evaluation and Reporting of Student Achievement. Washington: 1974, ED 099-405.
- Newell, A. Judgment and Its Representation: An Introduction. New York: Wiley, 1968. In B. Kleinmuntz (Ed.). Formal representation of human judgment.
- Newell, A., & Simon, H. Human Problem Solving. New Jersey: Prentice-Hall, 1972.
- Nie, N. et al. Statistical Package for the Social Sciences. New York: McGraw-Hill, 1976.
- Nisbett, R., & Ross, L. Human Inference: Strategies and Shortcomings of Social Judgment. New Jersey: Prentice-Hall, 1980.
- Pelto, P., & Pelto, G. Anthropological Research. New York: Cambridge Press, 1978 (1970).
- Phillips, B. Sex, social class and anxiety as sources of variation in school anxiety. Journal of Educational Psychology, 1962, 53, 316-322.

- Piliavin, I., Kodin, J., & Piliavin, J. Good semaritanisms: An underground phenomena? Journal of Personality and Social Psychology, 1969, 13, 289-299.
- Popham, W. Must all objectives be behavioral? Educational Leadership, April 1972, 29, 605-608.
- Prawat, R., & Nickerson, J. Affective outcomes in education. Progress Report, Institute for Research on Teaching, Michigan State University, 1979.
- Resnick, L., & Ford, W. The Psychology of Mathematics for Instruction. New Jersey: Erlbaum Associates, 1981.
- Rosenshine, B. Content, time and direct instruction. In P. Peterson and H. Walberg (Eds.). Research on Teaching: Concepts, Findings and Implications. California: McCutchan, 1979.
- Rosenthal, R., & Jacobson, L. Pygmalion in the Classroom: Teacher Expectation and Pupil's Intellectual Development. New York: Holt, Rinehart and Winston, 1968.
- Rugg, H. Teachers' marks and marking systems. Educational Administrative Supplement, 1915, 1, 117-142. Cited in Smith and Dobbins, Encyclopedia of Educational Research, 1957.
- Rudman, H. Integrating assessment with instruction, A review (1922-1980). Institute for Research on Teaching, Michigan State University, Research Series No. 75, January 1980.
- Ryan, K., & Levine, J. Impact of academic performance pattern on assigned grade and predicted performance. Journal of Educational Psychology, 1981, 73, No. 3, 386-392.
- Schatzman, L., & Strauss, A. Field Research, Strategies for a Natural Sociology. New Jersey: Prentice-Hall, 1973.
- Schellenberg, J. The class-hour economy. Harvard Educational Review, 1965, 35, 161-164.
- Schwab, J. The practical: A language for curriculum. School Review. Illinois: University of Chicago, 1969.
- Shane, J., & Shane, H. Ralph Tyler discusses behavioral objectives. Today's Education, September-October 1973, 62, 41-46.
- Shavelson, R. Research on teachers' pedagogical thoughts, judgments, decisions and behavior. Unpublished manuscript. University of California, Los Angeles, 1980.
- Shavelson, R., Caldwell, J., & Izu, T. Teachers' sensitivity to the reliability of information in making pedagogical decisions. American Educational Research Journal, Spring 1977, 14, No. 2, 83-97.

- Shulman, L., & Elstein, A. Studies of problem solving judgment and decision making. In F. Kerlinger (Ed.). Review of Research in Education. Illinois: F. E. Peacock, 1975.
- Simon, H. The Sciences of the Artificial. Cambridge: M.I.T. Press, 1969.
- Slovic, P., & Lichtenstein, S. Comparison of bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 6, 1971.
- Smith, A., & Dobbin, J. Marks and marking systems. Encyclopedia of Educational Research, (Ed.). MacMillan Co., 1959.
- Smith, E., & Sendelbach, N. Teacher intentions for science instruction and their antecedents in program materials. Paper presented at annual meeting of the American Educational Research Association, 1979.
- Smith, E., & Tyler, R. Adventures in American Education: Appraising and Recording Student Progress. New York: Harper and Bros., 1942.
- Stake, R. The case study method in social inquiry. Educational Researcher, A.E.R.A., February 1978, 7, No. 2.
- Starch, D., & Elliott, E. Reliability of the grading of high school work in English. School Review, 1912, 20, 442-457, idem. Reliability of grading work in mathematics. School Review, 1913, 21, 254-295; and Reliability of grading work in history. School Review, 1913, 21, 676-681.
- Stenhouse, L. Case study and case records; towards a contemporary history of education. British Educational Research Journal, 1978, 4, No. 2.
- Taba, H. Curriculum Development, Theory and Practice. New York: Harcourt, Brace and World, 1962.
- Taylor, P. How Teachers Plan Their Courses. Slough, Bucks: National Foundation for Educational Research, 1970.
- Thorndike, R. Marks and marking. Robert Ebel (Ed.). Encyclopedia of Educational Research, 4th Edition, MacMillan Co., 1969.
- Tiegs, E. Tests and Measurements for Teachers. Houghton, 1931. Cited in Smith and Dobbins, Encyclopedia of Educational Research, 1957.
- Traxler, A. Techniques of Guidance. New York: Harper, 1957.
- Traxler, A. The Nature and Use of Anecdotal Records. Educational Records Bureau, New York: 1949.
- Tyack, D. Ways of seeing: An Essay on the history of compulsory schooling. Harvard Educational Review, August 1976, 46, 355-389.
- Tyler, R. Basic Principles of Curriculum and Instruction. University of Chicago Press, 1950.

- Weiner, B. A theory of motivation for some classroom experiences. Journal of Educational Psychology, 1979, 71, No. 1.
- Weiner, B., Nierenberg, R., & Goldstein, M. Social learning (locus of control) versus attributional (causal stability) interpretations of expectancy success. Journal of Personality, 1976, 44, 52-68.
- Weinstein, M., & Fineberg, H., et al. Clinical Decision Analyses. Philadelphia: Saunders Co., 1980.
- Wilhelms, F. Evaluation as Feedback and Guide. Washington: Association of Supervision and Curriculum Development, Yearbook, 1967.
- Willis, S. Formation of teachers' expectations of students' academic performance. Unpublished doctoral dissertation, University of Texas at Austin, 1972. Cited by J. Brophy, in Teachers' cognitive activities and overt behaviors.
- Wisconsin State Committee. A uniform system of marking records in Wisconsin. Nations Schools, June 1929, 3, 50.
- Wrinkle, W. Improving Marking and Reporting Practices. New York: Holt, Rinehart and Winston, Inc., 1947.
- Yinger, R. A study of teacher planning: Description and theory development using ethnographic and information processing methods. Unpublished doctoral dissertation, Michigan State University, 1977. (A summary is available as Research Series No. 18. Institute for Research on Teaching, Michigan State University, 1978.)
- Zahorik, J. The effect of planning on teaching. Elementary School Journal, December 1970, 71, No. 3, 143-151.