DOCUMENT RESUME

ED 234 052 SP 023 019

AUTHOR Whitmer, Sylvia Pratt
TITLE A Descriptive Multimethod Study of Teacher Judgment during the Marking Process.
INSTITUTION Michigan State Univ., East Lansing. Inst. for Research on Teaching.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
REPORT NO IRT-RS-122
PUB DATE Jun 83
CONTRACT 400-81-0014
NOTE 37p.
AVAILABLE FROM Institute for Research on Teaching, College of Education, Michigan State University, 252 Erickson Hall, East Lansing, MI 48824 ($3.50).
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Elementary Education; Elementary School Students; *Elementary School Teachers; *Evaluation Methods; *Evaluative Thinking; *Grading; Student Behavior; Student Characteristics; *Student Evaluation; Teacher Attitudes; Teacher Behavior; Teacher Response

ABSTRACT

A study was designed to generate a description of 5 elementary school teachers' judgment processes during marking (of 152 students) across a school year. The findings support a model of the marking judgment constructed from the strategies and cues that emerged through analysis of marks, record books, and interviews. The model presents a three-phase process that was 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. The marking judgment is bounded by the classroom, a conclusion which suggests that many past marking studies have made assumptions about marks that are inappropriate to the teacher judgment process. The study found that formative marks serve as a feedback mechanism but that summative and final marks do not. Although specific conclusions are tentative because of the small sample size, the model is useful as a heuristic to generate further discussion, deliberation, and research hypotheses. Tables displaying report data are included as is an appendix listing teacher attribution-utility categories. (Author/JMK)

Research Series No. 122

A DESCRIPTIVE MULTIMETHOD STUDY OF TEACHER JUDGMENT DURING THE MARKING PROCESS

Sylvia Pratt Whitmer

Published By

The Institute for Research on Teaching
252 Erickson Hall
Michigan State University
East Lansing, MI 48824-1034

June 1983

Publication of this work is sponsored by the Institute for Research on Teaching, College of Education, Michigan State University. The Institute for Research on Teaching is funded primarily by the Program for Teaching and Instruction of the National Institute of Education, United States Department of Education. The opinions expressed in this publication do not necessarily reflect the position, policy, or endorsement of the National Institute of Eduation. (Contract No. 400-81-0014)

Institute for Research on Teaching

The **Institute for Research on Teaching** was founded at Michigan State University in 1976 by the National Institute of Education. Following a nationwide competition in 1981, the NIE awarded a second contract to the IRT, extending work through 1984. Funding is also received from other agencies and foundations for individual research projects.

The IRT conducts major research projects aimed at improving classroom teaching, including studies of classroom management strategies, student socialization, the diagnosis and remediation of reading difficulties, and teacher education. IRT researchers are also examining the teaching of specific school subjects such as reading, writing, general mathematics, and science, and are seeking to understand how factors outside the classroom affect teacher decision making.

Researchers from such diverse disciplines as educational psychology, anthropology, sociology, and philosophy cooperate in conducting IRT research. They join forces with public school teachers, who work at the IRT as half-time collaborators in research, helping to design and plan studies, collect data, analyze and interpret results, and disseminate findings.

The IRT publishes research reports, occasional papers, conference proceedings, a newsletter for practitioners, and lists and catalogs of IRT publications. For more information, to receive a list or catalog, and/or to be placed on the IRT mailing list to receive the newsletter, please write to the IRT Editor, Institute for Research on Teaching, 252 Erickson Hall, Michigan State University, East Lansing, Michigan 48824-1034.

Abstract

This study was intended to generate a description of the judgment processes of five elementary teachers during marking (of 152 students) across a school year. The findings support a model of the marking judgment constructed from the strategies and cues that emerged through analysis of marks, record books, and interviews. The model presents a three-phase process that was 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. The marking judgment is bounded by the classroom, a conclusion which suggests that many past marking studies have made assumptions about marks that are inappropriate to the teacher judgment process. The study found that 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.

A DESCRIPTIVE MULTIMETHOD STUDY OF TEACHER JUDGMENT DURING THE MARKING PROCESS[1]

Sylvia Pratt Whitmer[2]

The public's persistent dissatisfaction with teacher grading of student performance lies in a discrepancy between the functions ascribed to grades or marks by society and the functions actually taken into account by teachers when judging pupil performance in the classroom context. Society has used marks (1) as measures of academic achievement against an absolute standard (mastery), (2) as predictors of future achievement in grades K-12 (diagnosis and placement), (3) as predictors of college success (entry and credentialing), (4) as predictors of future job success (job entry and training), (5) as motivators for learning (reward and punishment), and (6) 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, 1976; Kirshenbaum, Simon, & Napier, 1971; Smith & Dobbins, 1959; Thorndike, 1969), 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; Joyce,

---

[1]This paper was presented at the annual meeting of the American Educational Research Association, New York City, March 1982. It summarizes (inclusive of key tables and figures) a doctoral dissertation, "A Descriptive Multimethod Study of Teacher Judgment During the Marking Process," College of Education, Michigan State University, December, 1981. Answers to methodological questions should be sought in the original document, which carries a detailed rationale for the methods used along with an extensive literature review and the five teacher case studies.

[2]Sylvia Pratt Whitmer, a former IRT research intern, is currently principal of Oakley Park Elementary School in Walled Lake, Michigan.

1979; Clark & Yinger, Note 1; Shavelson, Note 2). Immediate classroom demands and student characteristics heavily influence the marking judgment--the process of selection, organization and inference of evidence upon which the mark is determined. That is, teacher's 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 to the summary mark by society in general. The literature holds little on marking or teacher decision making in marking processes. This study attempts to determine the nature of the discrepancy and the teacher's mental process in mark selection.

Purpose of the Study

This study attempted to develop an understanding of the marking judgment that teachers engage in during the school year. Foremost was the goal to generate a description of the thoughts, judgments, and decisions of five elementary-school teachers during the marking task. In doing so I hoped (1) to identify strategies and cues that 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 that would indicate fruitful areas for future research.

Many highly involved constituencies--school districts, parents, teachers, students, and education researchers--have commissioned their own studies of marks, but have not targeted the teacher-judgment process. First, from the viewpoint of school districts and administrators, the report card remains the major communication device between schools and homes across the nation (Educational Research Service, Note 3). Parents rely on report cards as a personal pupil-progress report (Anderson, 1966). School districts and parents alike consider the marking process so important that district policies and teacher contracts specify periodic reports and often set aside paid teacher record days. Second, teachers view marking student work as a task that absorbs the most significant block of their professional time outside the classroom (Hilsum & Case, Note 4; Yinger, Note 5)

and that results in a rational system (record book) for explaining or justifying student marks at any time. Third, students view marks as part of a permanent record that may track them into specific skill levels or classes. Thus, marks continue to be the most reliable source of achievement information for determining eventual college or job entry (Bejar, 1981). Fourth, educational researchers view the process of marking from the perspective of its potential as the source of greatest teacher accountability in measuring student achievement. Yet, teacher-education programs seldom have courses or texts pertaining to the marking process or to its role within the larger teaching process.

Research Questions

Studying teacher-marking judgment is simply studying general human judgment. Judgment is well discussed by Johnson (1955) and Newell (1968) and summarized and reviewed by Shulman and Elstein (1975). The present study captured the marking-judgment processes of five teachers 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 that explains how the formative information is transformed into the summative mark?
4. If the judgmental rule yields a zone of uncertainty between any two preordained categories of judgment (A, B, C, D, or 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?

Methods

Four research strategies 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 teachers' methods of assessing the risk of their classroom behavior (decision tree); and attributional techniques to investigate teachers' methods of assessing risk related to future student motivation to achieve (interview data related to record book analysis and prediction data).

A multimethod approach to teachers' grading processes allowed the broadest description of the task. Using an integrated approach, I sought to maximize the strengths of each method while minimizing the weaknesses by carefully distinguishing the findings that several methodological perspectives corroborated from those that emerged in only one field of reference. In this manner, the study attempted to recreate teachers' understandings of the judgment task and to relate the task to achievement and management in each teacher's unique classroom.

Research Setting

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

Of the 28 school districts in Oakland County, pupil performance in District B is average. 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.

Following an initial expression of interest by five principals I contacted from schools using traditional marks, the first two contacts yielded five volunteer teachers--three men and two women--from grades 4-6. These five teachers became the subjects of the study. Experience beyond five years in the upper elementary classrooms was the only criterion I used for accepting teacher participants.

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

Procedures

Data Collection

The structured interview was the primary source of data acquisition. I interviewed and audiotaped each of the five participants on site immediately following the first, second, and final marking of the four periods in the school year. The tapes were subsequently transcribed into protocols.

The interviews were based on previous insights into interview formats and focused on products of the teachers' own creation, such as record books. This allowed teachers room for prediction, reflection, and open-ended responses.

I collected additional data from official marks, record books, and a pupil sort.[3] Marks of all students in each class included only language arts and mathematics, although the teachers also marked in spelling, reading, social studies, science, and art. I also asked teachers to predict the marks for each student for the next marking period and give brief reasons why they predicted that mark would remain the same, go up, or go down. The record-book data allowed a cross-check of teachers' verbal protocols.

Data Organization

The collected data were organized into a composite case and five individual cases. The composite case, described below, 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, appear in Whitmer (1981).

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.

Findings

The data originally collected on an individual basis, later became a composite model. The composite-case format served as an organizer, setting a pattern for

---

[3]Pupil sort: assignment of students to discrete categories such as "top of class," "above average," "below average," etc. based on effort and achievement.

describing the individual cases. The inferences are more extensive within the composite case. (See the Whitmer (1981) dissertation in which I show the basic data of each teacher case and make inferences. Discussion within each teacher case refers back to the composite case, noting points of difference.)

The composite case depicts the five teachers' commonalities (1) through computation of the marking data using multiple regression, correlations, and frequencies (National Institute of Education, 1980), and (2) through content analysis that 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 that guided the marking judgment of the five teachers: procedural and contingency. The rules dealt with different aspects of the judgment process. Procedural rules were concerned with selection and simplification of information being processed. These rules set up a linear, routine, record-book system; determined the tasks selected 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 for the five teachers determined judgment in uncertainty and exception. In the teachers' information processing, contingency rules concerned the inferential processes that went beyond the data. These rules essentially involved factors that 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 procedures and one of contingency judgment strategies.

Procedural rules:

1. The teachers assumed that completed tasks resulted in learning (implicit, not stated).
2. The teachers assigned tasks and gathered marking data regularly in a record book.
3. The 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. The teachers gathered marks from a sufficient variety of tasks (tests, written projects, exercises) to satisfy their criteria for validity. In any given marking period, no teacher had less than six formative marks. Four had more than 10.
5. The teachers had individual theories about weighting some tasks (tests vs. homework) more heavily than others.
6. The teachers had individual systems for transforming points representing standard criteria on a written paper into ABC marks.
7. The 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. The teachers ranked effort related to ability as a prime criteria for marking up or down. Effort was judged by regular work and extra work (record books and attribution chart).
2. The teachers had strategies to apply if the work fell midway between two marks.
3. The teachers had individual strategies for marks that fell below C (frequencies and quotations).

Procedural rules resulted in a record-book system that operated as a statistical tool to help overcome many of the common errors of human judgment, which are discussed in Nisbett & Roth (1980). An analysis of the record book showed the teachers' intent to account for a base rate of work (1) for the nation (the assignments adjusted to grade level on nationally normed, verbal information (i.e., textbooks)), (2) for the classroom (the vertical column of any given assignment), and (3) for the individual student (the horizontal row). Hence, the teachers' record books served as inferential tools depicting student achievement compared with 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, or E. However, when the work fell into a zone of uncertainty between two grades or when it fell into the D or E category, the teacher used the contingency rules. A statistical analysis of the marks (152 students) that resulted from the procedural rules follows.

Statistical Analysis

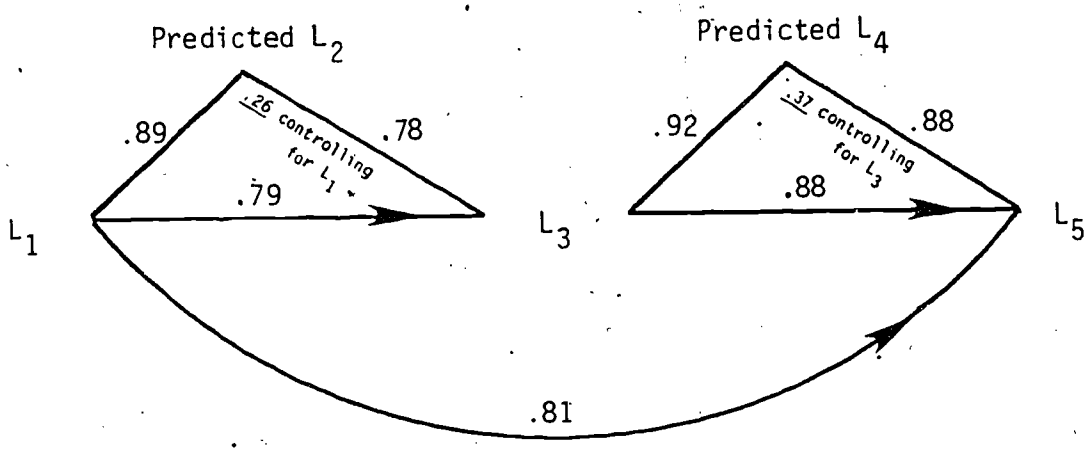
The statistical methods involved multiple regression analysis, P correlations, frequency counts, and cross tabulations.
computer for language arts and mathematics. The following symbols explain the marking data depicted in Figure 1. L_1 represents the first mark in language arts, L_2 the teacher's prediction of the second language arts mark, L_3 the second mark in language arts, L_4 the teacher's prediction of the final language art mark, L_5 the final mark in language arts. M_1 represents the first mark in math, M_2 the teacher's prediction of the second math mark, M_3 the second mark in math, M_4 the teacher's prediction of the final math mark, and M_5 the final mark in math.

For computation purposes, the summative marks, the predicted marks, and the final marks in both language arts and mathematics were assigned an arithmetical value and entered in the computer.

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

These values were used to derive all statistical factors found within the figures and tables of this paper. Their role is particularly told in the composite teacher-policy model (Figure 1). The judgment model is corroborated by the bar graph frequency pattern (Figure 2), illustrating that the average marks across the year are generally slightly lower than the teachers' predictions.

Language Arts

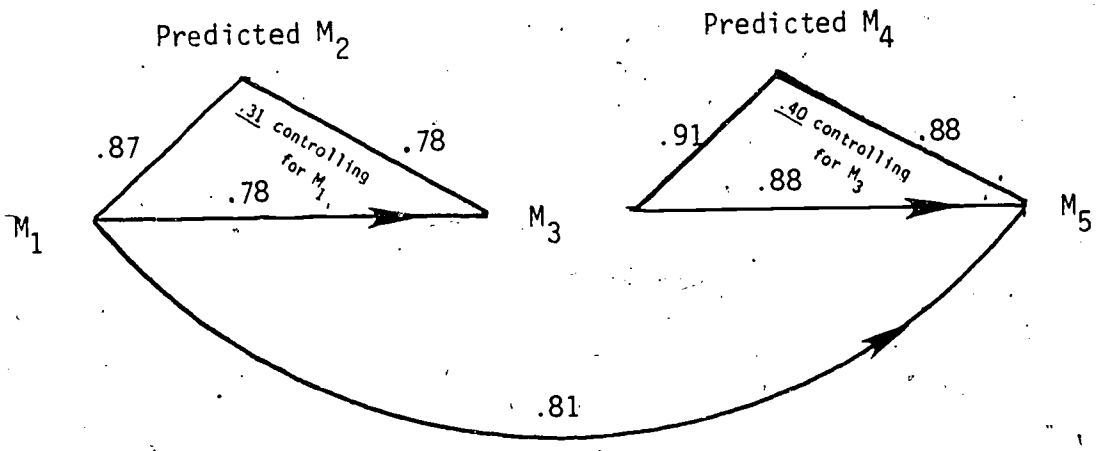


Note. L_1 = First actual mark

L_2 = Second prediction
L_3 = Second actual mark
L_4 = Final prediction
L_5 = Final mark

$r(L_4 L_5$ controlling for $L_1 L_3) = .24$

Mathematics



Note. M_1 = First actual mark

M_2 = Second prediction
M_3 = Second actual mark
M_4 = Final prediction
M_5 = Final mark

$r(M_4 M_5$ controlling for $M_1 M_3) = .32$

Figure 1. Marking policy[a] (with predictions) for all teachers

[a]These policies were captured through Pearson correlations adjusted by partial correlations. Summative marks and predicted marks of 152 students across a school year were the base data.

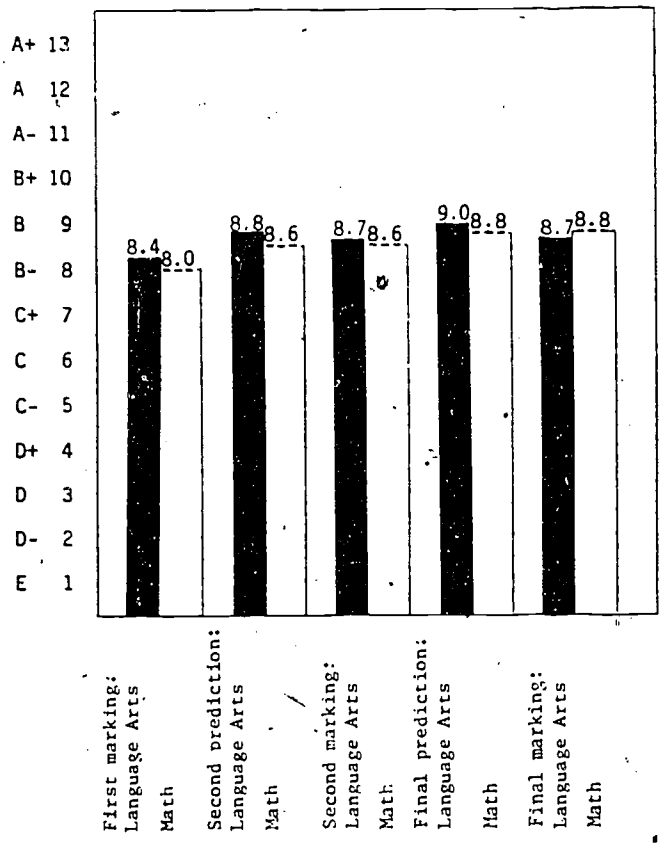


Figure 2. Composite pattern of marking averages across a year for all teachers (5) and all students (152).

Verbal Analysis

The marking rules that emerged through process tracing put the verbal analysis of protocols and the statistical analysis into perspective. This part of the study focused on the identification of the judgment factors underlying the contingency rules.

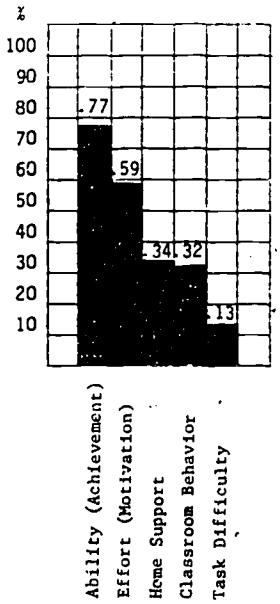
The teachers appeared to use contingency rules if they were uncertain about midway zones between marks and in cases of failure or near failure. Exposing the teachers' judgment cues involved various methods of establishing and categorizing teacher

concerns. In the interview process, I not only recorded the marks of 152 students, but asked teachers to predict the next marking and to discuss the factors that influenced their prediction.

Teachers marked students according to attributional categories of ability, effort, task difficulty, and luck. (See Appendix for elaboration of categories.) In coding verbatim responses, I used a miscellaneous category for one-time events. Early in the process an emergent "home-support" category replaced "luck" and an emergent "class behavior and physical maturity" category replaced the miscellaneous category. The latter is closely aligned with utility and maintenance of class flow or on-task behavior. Teacher statements were counted and percentages were determined (See Table 1).

Table 1
Composite Attribution-Utility

	First Marking				
	Teacher One 29 students	Teacher Two 28 students	Teacher Three 31 students	Teacher Four 33 students	Teacher Five 31 students
ABILITY (Achievement)	27/29	21/28	21/31	21/33	27/31
EFFORT (Motivation)	26/29	9/28	24/31	10/33	21/31
HOME SUPPORT	17/29	3/28	14/31	8/33	9/31
CLASSROOM BEHAVIOR + PHYSICAL MATURITY	7/29	15/28	6/31	12/33	9/31
TASK DIFFICULTY			10/31		9/31



Note. The left side of the table displays the actual count of attributions 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.

Class behavior, a utility concept, emerged as a category needing delineation. The teachers placed importance upon their ability to maintain on-task behavior and the flow of classroom activities. Maintenance of flow, a goal in itself, is a separate category from achievement but is related to it (Joyce, 1980). The teachers planned activities to accomplish academic tasks; they equated achievement with task completion. Therefore, any disruption of class flow took time away from a task. Individual students causing distractions lost time on task personally, but frequently, when a student disrupted the flow, everyone lost time on task. Where teachers perceived that sociability, excessive talking, and lack of concentration disrupted task-oriented behavior, they mentioned these characteristics in relation to predicted marks (e.g., "Her mark will probably go up when she controls her talking."). Each teacher stated that s/he allowed some level of conversation during class, hence, I interpreted any teacher comments on excessive talking, goofing off, teasing, and so on as off-task behavior that the teacher attempted to bring in line. Since the teachers based their marks on tasks completed, I assumed that when a teacher commented about a low grade s/he recognized that some students might get zeroes from incomplete tasks. Therefore, off-task behavior lowered a mark.

The category of classroom behavior lent itself to the decision-tree method of utility analysis (see Figure 3).

I found from this study that each marking period stood on its own tasks. The teachers did generally average formative marks at the end of a marking period, and did generally average the summative marks to arrive at a final mark for the year. However, an analysis of record books, of minuses and pluses, and of verbal protocols reveals that they did not do this as strictly or in as fixed a way as they perceived. Instead their contingency rules operated in zones of uncertainty and in exceptions. Contingency situations seemed to increase as the year went on.

The teachers shared common judgment cues in contingency zones. The cues included ability, effort, home-support level, classroom behavior/physical maturity, and task difficulty. Effort constituted the primary contingency cue, with ability close behind.

The composite study revealed that teacher-marking processes at the procedural level related to task completion, and those at the contingency level related to factors that promote task completion, especially effort. Interest in the home-support level basically related to gaining leverage to maintain or increase effort. Interest in classroom behavior 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.

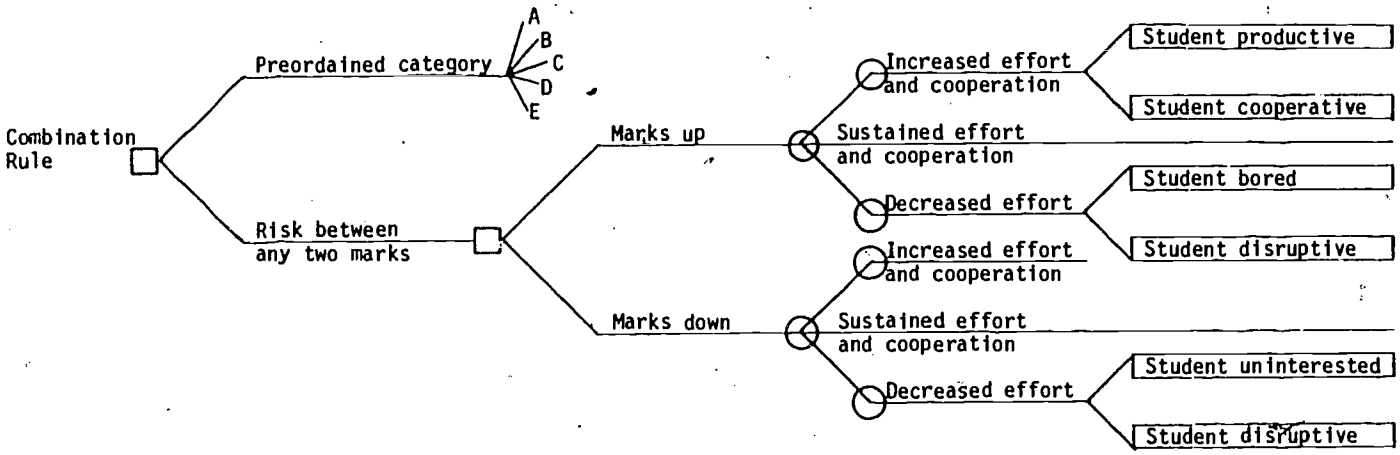


Figure 3. Decision tree: A utility framework for marking judgment. (Adapted from Weinstein, Fineberg, et. al., 1980, 18.)

Conclusions and Implications

Summary of Findings by Research Questions

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

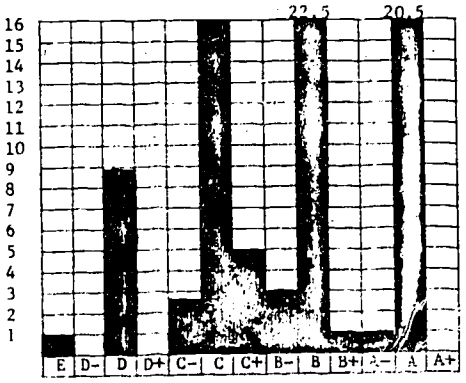
What cognitive processes make possible the formative stages (record-book categories) of marking? The cognitive processes of selection, simplification, and inference operate through heuristics (rules), attributions of individual success and failure, and perceived utilities[3] of the classroom. The record book was the key inferential tool of the process. Procedural rules emerged that guided and routinized it. The teachers varied in how they used 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 that explains how the input information (formative) is transformed into the output (summative) mark? Teachers used a linear arithmetic rule averaging across collected marks. This directly related to standard of mastery and degree of task completion. Within a marking period, this rule focused on completed tasks that carried weighted values and preordained categories of A, B, C, D, and E. For example, 10 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. Great discrepancy existed as to whether an E equals 0 or something above 0. Across the year, the rule focused on averaging the summative marks of each marking period. Hence the final mark was a derived arithmetic mean based on the weighted values of the completed tasks of each marking period.

[3]Utility is the measure of the usefulness of giving a particular mark or performing any activity. For example, if I give a child a B, will he work harder or not?

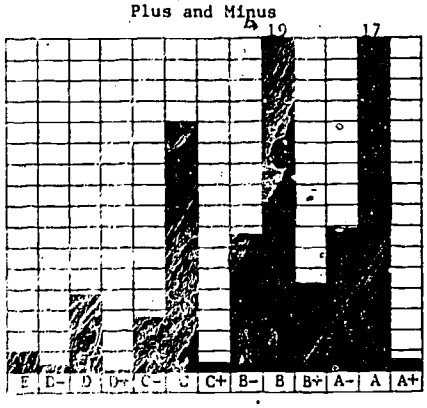
LANGUAGE

Marking Period = First (L_1)



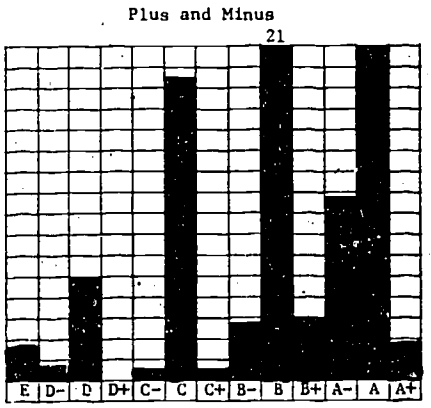
Note. % of minuses & pluses = 11.2
% of minuses = 7.9
% of pluses = 3.3
% related to A/B = 8.5
Preordained categories = 89.8

Marking Period = Second (L_3)



% of minuses & pluses = 27.5
% of minuses = 20.2
% of pluses = 7.3
% related to A/B = 23.8
Preordained categories = 72.5

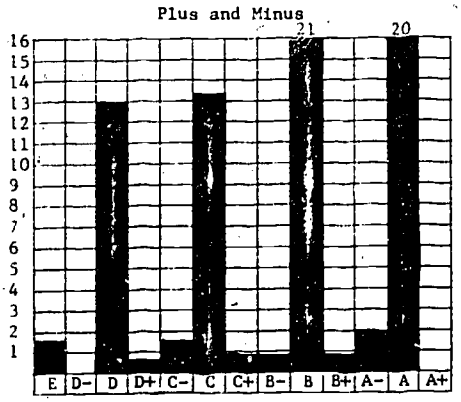
Marking Period = Final (L_5)



% of minuses & pluses = 22.5
% of minuses = 15.9
% of pluses = 6.6
% related to A/B = 20.4
Preordained categories = 77.5
Incomplete = 2
Blank = 2

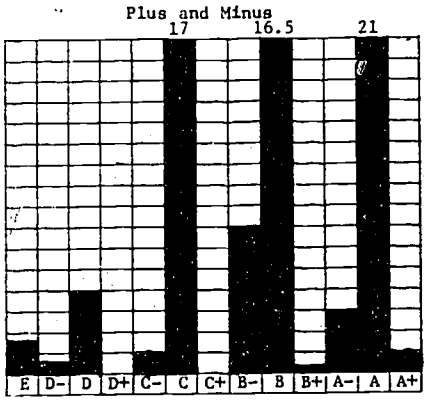
MATH

Marking Period = First (M_1)



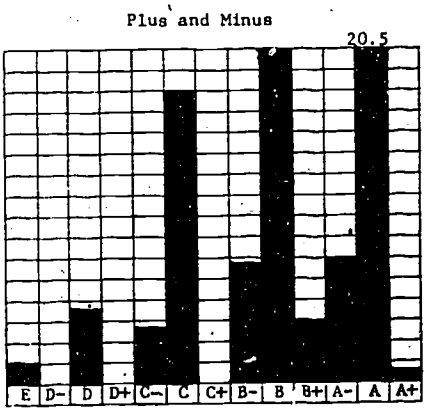
Note. % of minuses & pluses = 7.4
% of minuses = 5.3
% of pluses = 2.1
% related to A/B = 4.0
Preordained categories = 92.6
Incomplete = 0
Blank = 3

Marking Period = Second (M_3)



% of minuses & pluses = 17.1
% of minuses = 15.1
% of pluses = 2.0
% related to A/B = 15.1
Preordained categories = 82.9
Incomplete = 1
Blank = 5

Marking Period = Final (M_5)



% of minuses & pluses = 23.0
% of minuses = 18.4
% of pluses = 4.6
% related to A/B = 19.0
Preordained categories = 77
Blank = 5

Figure 4. Distribution of marks across three marking periods: A composite view of teachers' language and math marking.

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 that exceed 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.

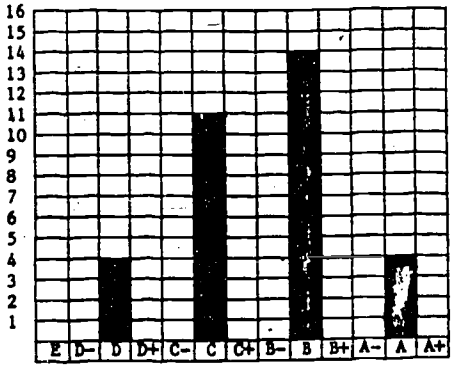
Do identified cognitive processes form a pattern, schema, or model of the marking process? A model was proposed. This model was based on the procedural and contingency rules that divided the marking process into three phases: selection and collection of data, valuing and assigning of data to preordained cateogires of A-E, and contingency factors to facilitiate choice under uncertainty or failure. The majority of marks were determined at the procedural level (See Figure 5).

Do identified cognitive processes account for the five functions ascribed to marks by society in general? I classified the functions into two general groups: One involved 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 involved conditions within the classroom structure such as motivation, achievement, and a teaching feedback function. I found that the judgment processes (rules, strategies, and cues) of the five teachers focused on task completion bounded by the particular classroom and its immediate participants. The marking-judgment processes of the

LANGUAGE

Marking Period = First (L_1)

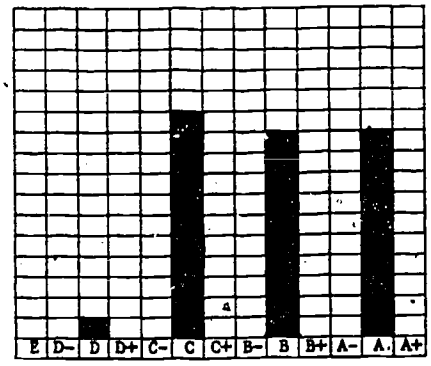
Plus and Minus



Note. % of minuses & pluses = 0
% of minuses = 0
% of pluses = 0
% related to A/B = 0
Preordained categories = 100%

Marking Period = Second (L_3)

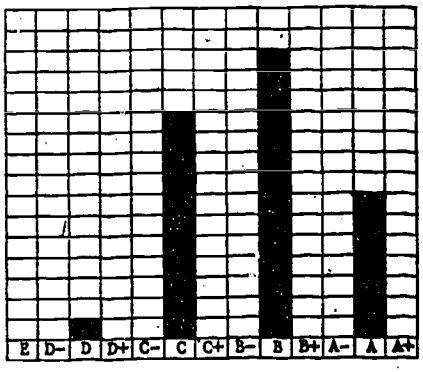
Plus and Minus



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

Marking Period = Final (L_5)

Plus and Minus

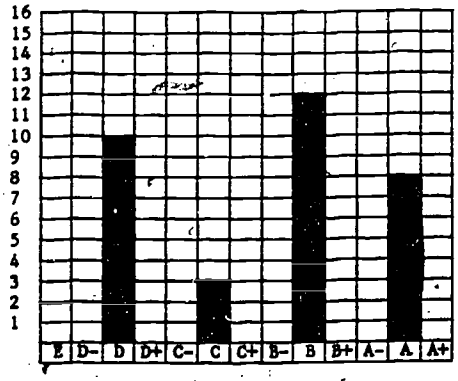


% of minuses & pluses = 0
% of minuses = 0
% of pluses = 0
% related to A/B = 0
Preordained categories = 100%
Blank = 1

MATH

Marking Period = First (M_1)

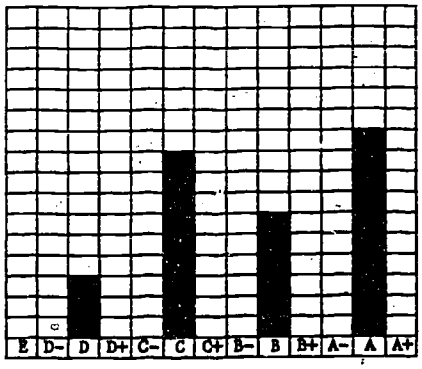
Plus and Minus



Note. % of minuses & pluses = 0
% of minuses = 0
% of pluses = 0
% related to A/B = 0
Preordained categories = 100%

Marking Period = Second (M_3)

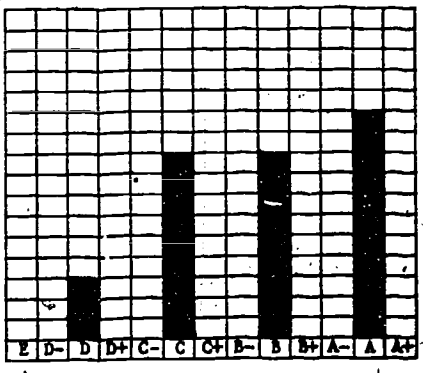
Plus and Minus



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

Marking Period = Final (M_5)

Plus and Minus



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

Figure 5. Distribution of marks across three marking periods for Teacher 4.

teachers, therefore, did not concern the functions ascribed to marks by those outside the classroom (school districts, parents, education researchers, etc.). Marking judgments primarily related to task completion at a given level of difficulty and standard of mastery, and to the factors promoting that completion. Hence the teachers defined their marking responsibility in terms of the practical demands of an average 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, (1) process tracing, (2) policy capturing, (3) attribution theory, and (4) utility theory shed light on different levels of the marking model. Process tracing allowed the broadest description of the marking judgment and supplied some part of the answer for each research question. Process tracing allowed many rules and cues used in the year-long marking process to surface. Based on a discussion of process training by Einhorn, Kleinmuntz, and Kleinmuntz (Note 6), a distinction between two subjudgment phases emerged for me. One dealt with choices between multiple categories (A-E). The other dealt primarily with a choice between any two categories. These phases, labeled procedural and contingency, 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[4] unless adjusted. Pearson correlations made a repeatedly strong case for a primacy effect. Partial correlations tended to adjust both techniques

[4]Recency is the tendency to weight one end of the marking process more heavily than the other because only one method of measurement has been used in past marking studies. Put another way, does the teacher tend to mark more heavily on papers at the end of the year, and do the resulting end marks more strongly affect the final grade than the first mark?

and supplied a modified policy that 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, once categorized and counted, illuminated the general weighting of the categories of ability, effort, home support, and task difficulty. Adjusted attribution charts show that effort counts more than ability but always vies with ability as the predominant criteria for marking judgment. This finding is substantiated by Weiner (1979) and discussed in Whitmer (1981). Policy capturing with statistical analysis did not get at these factors, but attribution theory with verbal analysis did. Frequency distributions of pluses, minuses, and checks further supported the findings, which showed that contigency 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, is concerned with contingency factors, particularly on-task behavior, with estimating of future effort or behavior, but not attributing cause on an individual basis. 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 the teachers' risks and thoughts when deciding to give a higher or lower grade.

Asking for a superior method was an inappropriate phrasing of the research question. Each method had its strengths and weaknesses. Together they provided a model for illustrating 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 judgments. Together they answered the research question about the five functions of marking, indicating that the validity of past research on marks must be questioned because it generally limits to single phases a much larger judgment process,

and it generally focuses on functions outside the classroom. Only with a multimethod approach was the total process illustrated (see Figure 6).

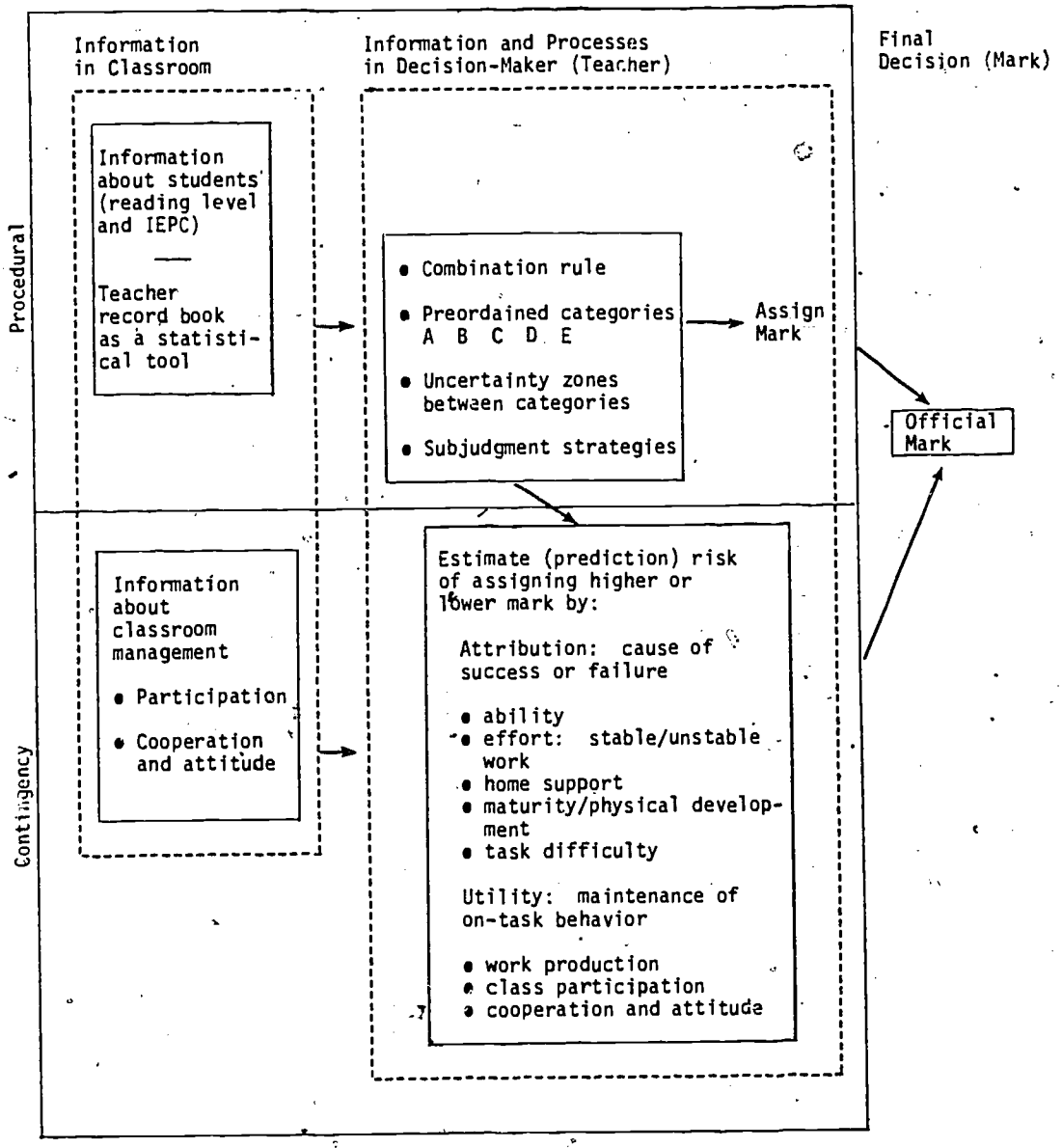


Figure 6. Framework for marking process (adapted from Carroll & Payne, 1976).

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's record book, appeared to carry a heavier weight than the quality of the completed work. Two features substantiate this assertion: Any work handed in received some credit above E. Students operating at lower-than-class average of task difficulty received the same amount of credit. Howev it is also notable that above the level of C, teachers began to create more categories of distinction by the use of minuses and pluses. Note the frequency distribution charts of marks across the year (Figure 5). Hence the criterion of completion had greater weight below C and the criterion of quality vied with completion above C. The criterion of completion was greater with students operating below grade level on task difficulty.

This emphasis on task completion at both an individual and class level calls into question the 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 appeared in this study 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 other studies. The marking judgment of teachers in this study relied directly on student task completion and indirectly on the classroom behavior that produced task completion more than it relied on identified student characteristics. This 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, Hoffer, & Kilgore 1981) or of the effective public schools (Brookover & Lezotte, Note 7) 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 that has implications for research is the bounded nature of the classroom and the fact that, actually, teachers think about and mark on events and interactions in the classroom. The linking of marks by society to events external to the classroom explains some of the previous unreliability of marks. The review of the marking literature indicates that many 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 that must be considered. The work of Doyle (1977, 1980), in particular, emphasizes the ecological nature of the classroom. The planning studies of both Yinger and Clark (Note 1; Note 8) 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 looks promising for future research. When tasks have been investigated in the past, only one task, such as a test on paper, has been examined. For example, the Starch and Elliott (1912, 1913a, 1913b) model of research asked a significant number of experts (100+) to correct one essay or test and concluded that marks were unreliable. My study suggests that the reliability of one task is discounted by the fact that the five elementary teachers collected 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 that the marking judgment model could play. In effect, it provides a framework for evaluating past marking studies, many of which were entirely involved with the procedural level of marking, others with the contingency level. Neither one alone accounts for the total marking process. Hence, the model places the value of past studies into a meaningful framework.

Implications for Practice

The heuristics of the study have implications for practitioners. Recalling Stenhouses's (1978) idea 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 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 teachers rely upon the home for leverage? To what extent do they communicate their procedural rules to the home versus being satisfied with the oft repeated combination rule statement that 90 to 100 is an A, 80 to 89 is a B, and so on, which is only a very small aspect of marking? In this regard, there may be obvious 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 teachers or researchers.

Finally, 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.

Reference Notes

1. Clark, C., & Yinger, R. Research on teacher thinking (Research Series No. 12). East Lansing, Michigan: Institute for Research on Teaching, Michigan State University, 1978.

2. Shavelson, R. Research on teachers' pedagogical thoughts, judgments, decisions and behavior. Unpublished manuscript, University of California, Los Angeles, 1980.

3. Educational Research Service. Reporting pupil progress: Policies, procedures and systems. Arlington, Virginia, 1977.

4. Hilsom, S., & Cane, B. The teacher's day. London: National Foundation for Educational Research in England and Wales, 1971.

5. 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, East Lansing, Michigan: Institute for Research on Teaching, Michigan State University, 1978.)

6. 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.

7. Brookover, W. & Lezotte, L. Changes in school characteristics coincident with changes in student achievement (Occasional Paper No. 17). East Lansing, Michigan: Institute for Research on Teaching , Michigan State University, 1979.

8. Clark, C., Yinger, R., & Wildfong, S. Identifying cues for use in studies of teacher judgment (Research Series No. 23). East Lansing, Michigan: Institute for Research on Teaching, Michigan State University, 1978.

References

Anderson, R. The importance and purpose of reporting. The National Elementary Principal, May 1966, 45 (6).

Bejar, I., & lew, E. Grade inflation and the validity of the scholastic aptitude test. American Educational Research Journal, Summer 1981, 18(2), 143-156.

Brophy, J. Recent research on teaching (Occasional Paper No. 40). Institute for Research on Teaching, Michigan State University, November, 1980.

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.

Doyle, W. Paradigms for research on teacher effectiveness. In L. Shulman (Ed.) Review 5, Illinois: Peacock, 1977.

oyle, W. Student mediating responses in teaching effectiveness: Final report (Project No..0-0645). Department of Education, North Texas University, March 1980.

Evans, F. What research says about grading. S. Simon & J. Ballanca (Eds.). Washington, D.C.: Association of Supervision and Curriculum Development, 1976.

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, 1980, 3, 73-77.

Kirschenbaum, H., Simon, S., & Napier, R. Wad-ja-get? The grading game in American education. New York: Hart, 1971.

National Institute of Education. Teaching as clinical information processing. Washington, D.C.: National Conference on Studies in Teaching (Report of Panel 6), 1975.

Newell, A., & Simon, H. Human problem solving. New Jersey: Prentice-Hall, 1972.

Nisbett, R., & Roth, L. Human inference: Strategies and shortcomings of social judgment. Englewood Cliffs, New Jersey: Prentice-Hall, 1980.

Shulman, L., & Elstein, A. Studies of problem solving judgment and decision making. In F. Kerlinger (Ed.), Review of research in education. Illinois: Peacock, 1975.

Smith, A., & Dobbin, J. (Eds.), Marks and marking systems. Encyclopedia of educational research (1st ed.), MacMillan, 1959.

Starch, D., & Elliot, E. Reliability of the grading of high school work in English. School Review, 1912, 20, 442-457.

Starch, D., & Elliott, E. Reliability of grading work in mathematics. School Review, 1913, 21, 254-295. a

Starch, D., & Elliott, E. Reliability of grading work in history. School Review, 1913, 21, 676-681. b

Stenhouse, L. Case study and case records; towards a contemporary history of education. British Educational Research Journal, 1978, 4(2).

Thorndike, R. In Robert Ebel (Ed.), Marks and marking. Encyclopedia of educational research, 4th Edition, MacMillan, 1969.

Weiner, B. A theory of motivation for some classroom experiences. Journal of Educational Psychology, 1979, 71 (1).

Whitmer, S.P. A descriptive multimethod study of teacher judgment during the marking process. Unpublished doctoral dissertation, College of Education, Michigan State University, 1981.

Appendix

Appendix

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, abnormally 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
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 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)
- Ill 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/Maturity/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 underdog, kind
Likes to please others
Likes to please me (teacher)
Yells out answers, lacks control
Nervous problems
Total emotional

Total behavior