NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

University of Alberta

Teachers' Beliefs and Philosophies About their Grading Practices

by

Elizabeth Ann Kushniruk 

A thesis

submitted to the Faculty of Graduate Studies and Research

in partial fulfillment of the requirements for the degree of

Master of Education

Department of Educational Psychology

Edmonton, Alberta

Fall 1994

ISBN 0-315-94944-9

Canada

Name Elizabeth Kushniruk

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

TESTS and Measurements
SUBJECT TERM

0288
SUBJECT CODE

U·M·I

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS
Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION
General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523

Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS
Language
- General 0679
- Ancient 0289
- Linguistics 0290
- Modern 0291

Literature
- General 0401
- Classical 0294
- Comparative 0295
- Medieval 0297
- Modern 0298
- African 0316
- American 0591
- Asian 0305
- Canadian (English) 0352
- Canadian (French) 0355
- English 0593
- Germanic 0311
- Latin American 0312
- Middle Eastern 0315
- Romance 0313
- Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY
Philosophy 0422
Religion
- General 0318
- Biblical Studies 0321
- Clergy 0319
- History of 0320
- Philosophy of 0322

Theology 0469

SOCIAL SCIENCES
American Studies 0323
Anthropology
- Archaeology 0324
- Cultural 0326
- Physical 0327

Business Administration
- General 0310
- Accounting 0272
- Banking 0770
- Management 0454
- Marketing 0338

Canadian Studies 0385
Economics
- General 0501
- Agricultural 0503
- Commerce-Business 0505
- Finance 0508
- History 0509
- Labor 0510
- Theory 0511

Folklore 0358
Geography 0366
Gerontology 0351
History
- General 0578
- Ancient 0579
- Medieval 0581
- Modern 0582
- Black 0328
- African 0331
- Asia, Australia and Oceania 0332
- Canadian 0334
- European 0335
- Latin American 0336
- Middle Eastern 0333
- United States 0337

History of Science 0585
Law 0398
Political Science
- General 0615
- International Law and Relations 0616
- Public Administration 0617

Recreation 0814
Social Work 0452
Sociology
- General 0626
- Criminology and Penology 0627
- Demography 0938
- Ethnic and Racial Studies 0631
- Individual and Family Studies 0628
- Industrial and Labor Relations 0629
- Public and Social Welfare 0630
- Social Structure and Development 0700
- Theory and Methods 0344

Transportation 0709
Urban and Regional Planning 0999
Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES
Agriculture
- General 0473
- Agronomy 0285
- Animal Culture and Nutrition 0475
- Animal Pathology 0476
- Food Science and Technology 0359
- Forestry and Wildlife 0478
- Plant Culture 0479
- Plant Pathology 0480
- Plant Physiology 0817
- Range Management 0777
- Wood Technology 0746

Biology
- General 0306
- Anatomy 0287
- Biostatistics 0308
- Botany 0309
- Cell 0379
- Ecology 0329
- Entomology 0353
- Genetics 0369
- Limnology 0793
- Microbiology 0410
- Molecular 0307
- Neuroscience 0317
- Oceanography 0416
- Physiology 0433
- Radiation 0821
- Veterinary Science 0778
- Zoology 0472

Biophysics
- General 0786
- Medical 0760

EARTH SCIENCES
Biogeochemistry 0425
Geochemistry 0996

Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES
Environmental Sciences 0768
Health Sciences
- General 0566
- Audiology 0300
- Chemotherapy 0992
- Dentistry 0567
- Education 0350
- Hospital Management 0769
- Human Development 0758
- Immunology 0982
- Medicine and Surgery 0564
- Mental Health 0347
- Nursing 0569
- Nutrition 0570
- Obstetrics and Gynecology 0380
- Occupational Health and Therapy 0354
- Ophthalmology 0381
- Pathology 0571
- Pharmacology 0419
- Pharmacy 0572
- Physical Therapy 0382
- Public Health 0573
- Radiology 0574
- Recreation 0575
- Speech Pathology 0460
- Toxicology 0383

Home Economics 0386

PHYSICAL SCIENCES
Pure Sciences
Chemistry
- General 0485
- Agricultural 0749
- Analytical 0486
- Biochemistry 0487
- Inorganic 0488
- Nuclear 0738
- Organic 0490
- Pharmaceutical 0491
- Physical 0494
- Polymer 0495
- Radiation 0754

Mathematics 0405
Physics
- General 0605
- Acoustics 0986
- Astronomy and Astrophysics 0606
- Atmospheric Science 0608
- Atomic 0748
- Electronics and Electricity 0607
- Elementary Particles and High Energy 0798
- Fluid and Plasma 0759
- Molecular 0609
- Nuclear 0610
- Optics 0752
- Radiation 0756
- Solid State 0611

Statistics 0463

Applied Sciences
Applied Mechanics 0346
Computer Science 0984

Engineering
- General 0537
- Aerospace 0538
- Agricultural 0539
- Automotive 0540
- Biomedical 0541
- Chemical 0542
- Civil 0543
- Electronics and Electrical 0544
- Heat and Thermodynamics 0348
- Hydraulic 0545
- Industrial 0546
- Marine 0547
- Materials Science 0794
- Mechanical 0548
- Metallurgy 0743
- Mining 0551
- Nuclear 0552
- Packaging 0549
- Petroleum 0765
- Sanitary and Municipal 0554
- System Science 0790

Geotechnology 0428
Operations Research 0796
Plastics Technology 0795
Textile Technology 0994

PSYCHOLOGY
General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

University of Alberta

Release Form

Name of Author:	Elizabeth Ann Kushniruk
Title of Thesis:	**Teachers' Beliefs and Philosophies About Their Grading Practices**
Degree:	Master of Education
Year This Degree Granted:	1994

E.A. Kushniruk
(Student's signature)

16536 - 113 Street
Edmonton, Alberta
T5X 3V7

Date Sept. 28, 1994

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled "Teachers' Beliefs and Philosophies About Their Grading Practices" submitted by Elizabeth Ann Kushniruk in partial fulfillment of the requirements for the degree of Master of Education.

Dr. W. Todd Rogers, Supervisor

Dr. Thomas O. Maguire

Dr. Joyce Edwards

Date Sept. 26, 1994

To my beautiful children, Karla and Karin,

who by their efforts consistently

'make the grade' and make me so proud

Abstract

The purpose of this study was to investigate teachers' beliefs and philosophies of their grading practices and to identify potential areas of deficiencies or discrepancies. The subjects were 19 teachers in a large urban school district who had previously taught or were currently teaching Grades 4-6 students. The instrument was a semi-structured interview developed from prior research, measurement theory, and practical teaching experience. It was reviewed and pilot tested with a small sample, and the main research was conducted in April and May of 1994.

The in-depth interviews were transcribed and processed using qualitative data analysis. The teachers' protocols were categorized into three main topics for the presentation of the findings and discussion: functions of grades, processes of grading, and contextual factors which served as a mediatory role in teachers' grading practices.

The results showed that the teachers had a good understanding of the functions and processes of grading but that contextual factors often modified their grading practices. Their beliefs generally prompted them to seek a practice that was beneficial for an individual student at the risk of sacrificing a common meaning of the grade. In light of contextual factors, both internal and external to the student, and the freedom of teachers to practice their beliefs, the validity of a grade may become questionable.

Some of the main discrepancies which lead to questions and which were identified from the teachers' responses were: What student characteristics should be included in grades? How should students of differing abilities be graded? and Should academic and nonacademic subjects be graded differently? Differences among teachers were also noted in the way they aggregated scores prior to assigning grades and in their self-proclaimed use of inconsistent and subjective decision making about grades.

However, all the teachers believed that consistency in grading practices was essential to the accuracy of grade interpretation and suggested adopting either a school-wide grading policy or a common plan for each division within a school. Further, the teachers realized that the implementation of such a policy would be difficult. They appealed for more direction from the district's central office in terms of standardization of measures and procedures. Implications for practice and research were suggested.

Acknowledgements

The writer wishes to gratefully acknowledge the people who assisted in the completion of this thesis. Sincere appreciate is extended:

To Dr. Todd Rogers, my advisor and mentor, who provided me with encouragement, advice, and his precious time to read the endless drafts and figures, as I tried to decipher the meaning of this work. His penchant for "scholarly work" can only be admired.

To Dr. Tom Maguire and Dr. Joyce Edwards for their contributions as committee members.

To the volunteers of this study. I was fortunate to have had the opportunity to meet 19 dedicated teachers who took time out of their busy schedules and so graciously allowed me to interview them. I am appreciative of the honest and candid manner in which they responded to my questions.

Table of Contents

List of Tables

List of Figures

Chapter I

The Research Question

Introduction

In our daily lives we evaluate with astonishing frequency. Evaluative terms form one of the basic dimensions of meaning in our language (Terwilliger, 1971) and are often defined by pairs of adjectives—for example, pass-fail, excellent-poor, strong-weak, high-low—that act as 'sorting' devices. In a society which places great emphasis on individual achievement, evaluation encompasses a differentiation or a form of sorting of individuals in an effort to acknowledge their talents and skills. According to Terwilliger, "It is primarily the teacher who transmits the knowledge, skills, and values of [our] society and it is also the teacher who judges the extent to which these have been acquired by students" (p. 5).

Because a great number of value judgments regarding students are made by teachers, grading becomes a vital issue in the lives of both students and teachers. Grades serve as one of the most powerful tools for communicating with students (Manke & Loyd, 1991) and parents (Schulz, 1993; Waltman & Frisbie, 1993), are a source of affirmation and guidance to students (Barnes, 1985), and form an important component of decision making for teachers, parents, and students (Linn, 1990). A grading policy can be "a lever for establishing a schoolwide standard to which students and teachers aspire and a tool for motivating students to reach that standard" (Wiggins, 1988, p. 20).

Gronlund and Linn (1990) described the reporting of the summary results of various evaluative activities as "one of the more frustrating aspects of teaching . . . [with] so many factors to consider . . . and so many decisions to be made" (p. 427). Terwilliger (1971) has described assigning grades as "probably the most unpopular task that a classroom teacher must perform" (p. 7). The difficulty of functioning as both judge and advocate of the student at the same time has been raised (Brookhart,

1992). In addition to the signals that grades convey to students is the burden of knowing that grades "portray the student to others who may take the grades as valid evaluations" (Leiter & Brown, 1983). Because teachers recognize that evaluation involves judgments which may have far-reaching consequences, it is "rarely frivolous" (Barnes, 1985, p. 46).

Grading has not received the same attention as other facets of teacher education such as teaching methods and classroom management have (Allal, 1988; Manke & Loyd, 1991); however, other facets of evaluation, in particular, testing, have been well established (e.g., Dorr-Bremme, 1983; Griswold, 1988; Gullickson, 1984, 1985; Marso & Pigge, 1992). The recent renewed interest in evaluation has been focused on classroom testing and performance assessment (Stiggins, 1988; Stiggins & Bridgeford, 1985). Though grading guides have been written for teachers (Frisbie & Waltman, 1992; *Principles,* 1993; Stiggins, 1991; Terwilliger, 1977, 1989; Wiggins, 1988) and recently written books have been added to professional libraries to supplement the existing plethora of measurement books (Ebel & Frisbie, 1991; Gronlund & Linn, 1990; Mehrens & Lehmann, 1991; Oosterhof, 1990; Popham, 1990), the grading of students continues to be a highly controversial and problematic practice. Recent work, however, has indicated a move towards the identification of how teachers view grades and what they hope to communicate with them (Schulz, 1993).

Debates on Grading

In addition to grading being a "distasteful" (Terwilliger, 1977, p. 1) aspect of teaching, it is also one topic in education which is subject to heated debates. The grading process has become a "complex, varied, subjective procedure, thereby eliciting questions on its continued practice" (Nava, Josefa, & Loyd, 1992, p. 21). The abolition of the grading system has been expressed (Nava et al., 1992;

Terwilliger, 1977), and recommendations for alternative practices have been established (Burton, 1983; Crooks, 1988; Simon & Bellanca, 1976; Stiggins & Conklin, 1992). However, Ebel (1974), in summarizing the arguments over grading, presented a strong case in favor of grades.

The formation of grades is a highly subjective process. Teachers may use different standards, attach different purposes or importance to the grading process, use different criteria in determining grades, or assign grades according to "preferences for certain student attitudes and behaviors; . . . [for example,] compliance and involvement" (Austin & McCann, 1992, p. 2). Questions concerning the reliability (Stiggins & Bridgeford, 1985; Thorndike & Hagen, 1977) and the validity (Griswold & Griswold, 1992; Manke & Loyd, 1991) of grades have been raised, as have concerns about the lack of teachers' formal training (Gullickson, 1986) and of the quality of their assessments (Stiggins & Bridgeford, 1985) from which grades ultimately are derived. Other studies, however, have shown teacher judgment to be accurate (Hoge & Coladarci, 1989).

Grading practices have been the subject of criticism. They may easily become a contributing factor to gender or ethnic biases (Griswold & Griswold, 1992) or to social desirability, especially if based on "structured and spontaneous performance assessments" (Stiggins & Bridgeford, 1985, p. 283). According to Nava et al. (1992), the objectivity and credibility of the grading process is being challenged because "little is known about the specific criteria that teachers, in general, include in grading, or how the specific criteria are evaluated and used in combination when teachers make decisions on a student's end-of-term grade" (p. 21).

The meaning (Brookhart, 1992; Griswold & Griswold, 1992) and purpose (Austin & McCann, 1992) of grades are often varied. Variability and inconsistencies in grading reflect the complexities inherent in the purpose of schooling. As important as knowledge and skills is the socialization of students for their successful future

participation in our society. Teachers have the task of producing responsible, co-operative, and hard-working citizens. The traditional idea of grades reflecting only academic achievement has, therefore, been challenged. Indeed, as pointed out by Griswold and Griswold (1992), grades often include other social factors. Further confounding the meaning of grades

> are the needs of different stakeholders . . . for evaluative information in an economical form. They want, to varying degrees, information about the status of current achievement, about effort being made and development occurring, and about standing in comparison to the peer group. (Austin & McCann, 1992, p. 12)

Consequently, the validity of grades becomes questionable as teachers' perceptions of grades and the users' interpretations differ (Anderson & Bachor, 1993; Pilcher-Carlton & Oosterhof, 1993; Waltman & Frisbie, 1993).

Teachers do not follow recommended grading practices (Stiggins, Frisbie, & Griswold, 1989). Gronlund and Linn (1990), Hopkins, Stanley, and Hopkins (1990), Popham (1990), and Terwilliger (1977) all supported basing grades solely on demonstrated achievement, unclouded by teachers' subjectivity. Yet over 50 years ago Scates (1943) observed fundamental distinctions between the criteria which measurement experts use and the criteria that governs most teachers' practice. He described teachers' strategies as being based on "interplay between objectivity and judgment" (p. 6).

Factors other than achievement have been included in the formation of grades (Allal, 1988; Austin & McCann, 1992; Manke & Loyd, 1991; Nava et al., 1992; Stiggins et al., 1989; Wood, Benner, & Wood, 1990). Furthermore, studies by Brookhart (1992) and Griswold and Griswold (1992) and guidelines written for teachers in their assessment of students (*Principles,* 1993) indicated that teachers did not intend grades to reflect only achievement. The use of nonachievement criteria has been justified because achievement is affected by such variables (Brookhart, 1992;

Nava et al., 1992), and it is believed that they should be measured (Manke & Loyd, 1991; Schulz, 1993).

It has been suggested that measurement specialists' view of grading does not match the complex realities of the classroom (Brookhart, 1992; Stiggins et al., 1989). There are "logical and situational reasons for different teachers to distribute grades differently—and to weight different factors in their grading" (Wood et al., 1990, p. 10). Ability of students (Manke & Loyd, 1991; Stiggins et al., 1989; Wood et al., 1990) and consequences for students (Brookhart, 1992; Pilcher-Carlton & Oosterhof, 1993) are factors considered by teachers.

According to Stiggins (1988), the relationship between grades and student motivation is not completely understood. Furthermore, the question of grades as motivators has been debated. Although Brookhart (1993) suggested that grades are used for motivation and class management, other research findings (Crooks, 1988) claimed that motivational influences are equivocal and that students' responses to evaluation are individualistic and highly contextual.

In addition to the broad issues discussed above, specific ones have been addressed in studies. For example, there is variability in what types of measures are used to form grades (Anderson & Bachor, 1993; Nava et al., 1992; Stiggins & Bridgeford, 1985; Wood et al., 1990), how scores from assessment measures are combined to obtain summative grades (Allal, 1988), and how grades are distributed across students in different programs (Bateson, 1990; Wood et al., 1990). Although it is customary to apply different standards to students who are provided with different objectives—for example, special-needs students—Wood et al. (1990) suggested that differences in grading practices "should be especially obvious in a comparison of teachers of students who vary most in initial abilities [such as] athletes [or] musicians" (p. 6).

Stiggins et al. (1989) reported that an examination of practices that teachers use to grade student performance is not found in the recent literature. More specifically, what is missing is

> an analysis of the underlying assumptions and philosophies teachers use in the grading process. Also missing is a summary of the actual practices teachers use to generate grades: the student characteristics they use, the measurement procedures they use, their rules of evidence, or the standards they apply. (p. 6)

Though studies about what teachers do in their actual grading practices have been conducted (Allal, 1988; Bateson, 1990; Leiter & Brown, 1983; Stiggins et al., 1989; Wood et al., 1990; and others), there have been "only a few studies [which] have asked teachers what they think and feel about this part of their work" (Manke & Loyd, 1991, p. 3). According to Frary, Cross, and Weber (1992), in those studies in which teachers' grading practices were documented "none . . . attempted to measure individual teachers' beliefs in order to identify areas of deficiency for possible remediation" (p. 1).

Purpose of the Study

The main purpose of this study was to investigate the beliefs and feelings that teachers have concerning the grading of upper-elementary-school-aged (Grades 4-6) students. The researcher's task was to elicit from teachers' responses a number of selected issues related to grading practices identified in the literature. In the tradition of ethnographic research, the study was aimed at understanding the viewpoints, perceptions, and beliefs of a group of teachers about the particular phenomenon of grading. The constructed realities (Lincoln & Guba, 1985) formed from these perceptions and belief systems were reconstructed during data analysis. A second purpose was to identify discrepancies within teachers' grading practices and to discuss implications for the further grading practice of teachers.

Significance of the Study

Of foremost interest are the references in the literature pointing to a lack of research on the subject of grading. According to Manke and Loyd (1991), the beliefs of practising teachers about grading have been little studied. A particular observation is that, with the exception of the research of Thiessen and Moorehead (1985), there have been no Canadian studies of this nature located in the literature. According to Henry Schulz (personal communication, March 1994), research related to the issues of grading is a much-needed effort in Canada.

This study is unique in that by applying parts of existing research to a new geographical setting, a comparison between the beliefs of teachers in the United States and elsewhere and their Canadian colleagues can be made. Because the majority of the extant literature on grading relates to high school- or college-aged students, some of the issues that relate to the secondary level are being examined in the event that insights can be provided at the elementary level. Although it is commonly accepted that grading and reporting practices of elementary and secondary school teachers differ, this study provides an opportunity for upper-elementary teachers to focus on their beliefs in the knowledge that beyond Grade 6 their students may be exposed to a much different grading system.

The topic of grading is part of a larger movement, within educational circles, of re-defining student evaluation. This study was timely in that its inception coincided with a refocus and new direction of grading and reporting student progress (Dale Armstrong, personal communication, November 1993). It is hoped that by investigating teachers' beliefs about grading, new information may be brought to bear upon some of the existing weaknesses in the grading forum.

This study investigates a wide range of issues related to grading. The topics range from the meaning and influences of grades, identification of student characteristics and frameworks to use when forming grades, and consistency of

grading practices, to specific issues such as identification of measures and symbols to use, grade-assignment procedures, and elaboration of grading difficulties experienced by teachers. The goal of identifying potential discrepancies in order to inform various educational groups was made possible by the breadth of this topic. The results of this study have implications for the future practice of teachers.

Nature of the Study

As a branch of qualitative research, educational ethnography characteristically "provides detailed descriptions about the contexts, activities, and beliefs of selected participants in educational activities" (McMillan & Schumacher, 1989, p. 388). As previously indicated, the main purpose of this study was to describe and interpret those beliefs pertinent to teachers' grading activities. Though based on a discovery orientation and an emergent design, the approach taken in this study, with its use of a semi-structured interview, is best supported by an eclectic approach in terms of data collection and analytic strategies.

Definition of Terms

The following definitions were used to provide consistency of meaning throughout the study.

Grade: A symbol or mark (i.e., number, letter, word) which represents a value judgment made by a teacher concerning the relative quality of a student's achievement of learning objectives during a specified period of instruction; a mark. Most often thought of as a percentage or a letter, a grade also includes levels—for example, at, above, or below grade level—or categories—for example, E, S, or N (excellent, satisfactory, needs improvement). A grade is a 'shorthand' language for communicating evaluative information about students. The meaningfulness of grades

depends on the extent to which a school/community has a shared understanding of what they represent.

Grading: The process of assigning grades. The 'value' placed on a student's work may be relative to (a) others (one student achieved an A compared to another who achieved a C); (b) a standard (a student has achieved 80% or mastery of a topic); and (c) him/herself (a student has shown an improvement in sentence structure, and therefore there is evidence of personal growth). To be excluded from grading is a pure anecdotal form of description—for example, stating what a child can or cannot do—where there is no reference to external criteria).

Growth: Evidence of positive change in student achievement or attitude over time; improvement as a result of learning.

Achievement: A variable of the quality of performance in the subject matter only. Achievement can be stated in terms of the product (e.g., test, project, essay) or process (e.g., observation of a debate, skill in conducting an experiment).

Nonachievement: Variables other than achievement which include effort, work habits, attitude, potential for learning (aptitude), personality, and citizenship (e.g., cooperativeness).

Preset distribution: A method of basing grades on the normal curve. The distribution is determined prior to collection of scores. For example, a teacher may choose to use the following distribution of grades: 10% get A, 15% get B, 55% get C, 15% get D, and 5% get F.

Preset standard: A method of basing grades on a defined standard. It may be a percentage denoting mastery or a level based on achievement of specific learning objectives.

Empirically derived grades: Grades are determined by the teacher after the scoring is complete.

Evaluation: The systematic process of collecting, analyzing, and interpreting information to determine the extent to which pupils are achieving instructional objectives.

Performance assessment: The observation and rating of student behavior and products in contexts where students actually demonstrate proficiency.

Assumptions

The basic assumption underlying this study was that the teachers' comments and responses would accurately reflect their beliefs and philosophies of grading. Second, it was assumed that the responses were not biased for those respondents known to the researcher.

Delimitations of the Study

The study was restricted to teachers of upper-elementary-aged students (Grades 4-6) who taught in a large urban school system.

Organization of the Thesis

A brief history of grading, followed by a review of the studies relevant to the purpose of the present study, is presented in Chapter II. In Chapter III the research design and procedures are described. Chapters IV, V, and VI provide detailed findings of the study. A summary of the findings, including an integration of the literature; conclusions; and implications for further practice and research are presented in the seventh chapter.

Chapter II

Literature Review

Introduction

Stiggins, Frisbie, and Griswold (1989) reported that "a review of the grading research literature yields little that is of value in answering" (p. 6) grade-related questions. Furthermore, the majority of the studies located in the literature have been directed toward the grading practices of high schools or postsecondary institutions and conducted outside Canada. Few studies have focused directly on investigating the beliefs that teachers have about grading or how they think as they work through the process of grading, and fewer still have aimed at interviewing teachers of elementary students.

Prior to a discussion of the literature, a brief history of grades is presented. Following the history are the studies selected for this review, which have been divided into three sections. In the first section those studies most similar in their content and focus to that of the present study are presented and reviewed. In the second section a general review is presented. This review includes those studies related to the present study in terms of their main findings, but where the focus or approach differs from the present study. These studies are grouped according to similarity of topic. Studies conducted in Canada are presented in the third section.

A Brief History of Grading

Throughout history the forms of evaluation have changed, but the fact of evaluation has not (Terwilliger, 1971). Although the concept of 'sorting' is age-old, grading is a relatively recent phenomenon (Hargis, 1990). Its widespread use coincided with the advent of public education in the United States. Prior to 1850 progress was indicated descriptively; for example, when a student had acquired the correct number of skills, he was ready for the next level of study. As the number of

public high schools increased dramatically between 1870 and 1910 (Kirschenbaum, Simon, & Napier, 1971) and more grade levels and subject areas were introduced into the curriculum (Hargis, 1990), percentage scales or scales of 1-100 began to be used as a way of differentiating students of various academic abilities. This system served administrative functions in the organization of schools and in the screening processes for higher education. According to Hargis, grades had no educational purpose other than to classify students.

In time the percentage and 1-100 scales came under scrutiny. In the classic 1912 study by Starch and Elliot (cited in Hargis, 1990), it was revealed that personal values and expectations of teachers vastly influenced their grading standards, even in subjects such as math. To improve objectivity and reliability (the authors felt that teachers could not reliably assign grades in less than a 7-point error band), letter grades or point scales were adopted. To overcome grade inflation or harsh graders, grading on the normal curve became a popular practice in a further attempt to objectify and add "rigour to evaluation" (Hargis, 1990, p. 15). Though several distributions based on the curve were presented, with the most common being the 7-24-38-24-7 distribution designed by Cajori in 1914 (Hargis, 1990), the actual practice of grading on the curve is not widely used (Hargis, 1990).

With the practice of grading on the curve came the realization that most classes did not contain a cross-section of the population, that ability was not normally distributed. Hence, grading according to ability became popular, and as the number of intelligence and standardized tests increased dramatically in the first decades of the twentieth century, the practice of distributing grades based on ability grouping became widely used (Hargis, 1990). However, the reliability of teachers' grading practices did not improve. In studies regarding the variability and reliability of grades conducted in the 1950s, "considerable discrepancy between grades teachers gave for achievement and scores on standardized achievement tests" (p. 19) were noted.

The progressive education movement of the 1930s and other humanistic trends sought the abolition of formal grading (Hargis, 1990). There was an emphasis on mastery learning approaches which accommodated individualized student progress. Some school districts adopted pass-fail systems, in which the only important division was between acceptable and failing work.

During the 1940s grading continued to be contentious in nature, with the measurement camp and the progressives highly polarized (Kirschenbaum et al., 1971). The abolitionists felt that grading was unduly competitive and would have a negative effect. They saw descriptive evaluation serving communication and thus meaning better. Other critics found less fault with the grading system, although concerns remained about the misuse of the grading system and the lack of clear objectives and marking criteria. These people argued that there would always be a need for some system to compare individuals' achievements. By the end of the 1940s, 80% of the nation's schools employed some form of the 5-point (A, B, C, D, F) scale (Kirschenbaum et al., 1971). Though there was still talk about "normal curves, objectivity, specifying grading criteria . . . problems of reliability and validity, superficial descriptions, grade competition, and damaged self-concepts" (p. 68), there was little impact on or change in the letter-grade system.

According to Hargis (1990), the systems of grading that were adopted by the end of the 1920s are still much in use today. Interest in and use of a particular system are largely a function of the "ebb and flow" (p. 19) of various trends in education. New developments in measurement—for example, ratings of performance assessments—have produced a greater variety of descriptors to signify levels of performance.

Today many teachers' grades are based on a standard scale, sometimes mandated on a systemwide basis or left to the policy of individual schools. Often scores are compared to a scale and grades are determined without any regard to a

distribution of scores (Hargis, 1990). Furthermore, cut-off points for each grade vary widely (Hargis, 1990; Stiggins et al., 1989). According to Frary et al. (1992), many teachers perceive percentage-grading scales as representing absolute measures, although it is widely known that many scores from tests provide no more than ranking information.

A survey conducted by Kunder and Porwoll in 1977 (cited in Wood et al., 1990) showed that letter grades were used by 90% of senior high schools, 91% of junior high schools, 69% of upper elementary grades, 37% of primary grades, and 8% of kindergarten teachers. According to Hills (1981) and Friedman and Frisbie (1993), letter grades are still the most popular symbol for regular school subjects. Although Kunder and Porwoll reported that 70% of elementary schools used parent-teacher conferences as a form of communication of student progress, according to Schulz (1993), parent-teacher conferences appear to be common in most Canadian schools.

Teachers' Beliefs and Thought Patterns About Grading

The following five reviews—Manke and Loyd (1991), Griswold and Griswold (1992), Brookhart (1992, 1993), and Barnes (1985)—relate specifically to studies in which teachers' beliefs and thoughts about grades were described.

The purpose of Manke and Loyd's (1991) study was to explore what teachers thought was important about grading and what troubled them about the grading processes they used. By recognizing those teachers' current needs and beliefs about grading, it was hoped to identify the emphasis that grading should receive within teacher-education programs in order to bridge the gap between teachers' measurement knowledge and the practical needs of the classroom.

The subjects were 105 teachers who supervised student teachers enrolled at a university in the southeastern United States. The respondents were asked to identify

the grade level they taught and the kinds of grades they used on their report cards. They were then asked to describe their philosophy or basic approach to grading and the difficulties that they encountered in carrying out this process.

Using a content analysis, the results of the third question were clustered into themes. Because of the nature of the sample, the results were divided into four groups: secondary, upper elementary, primary, and special education. The authors found two dominant themes among third- through sixth-grade teachers. The first was that grades were important as a means of communication. Teachers felt that grades provided opportunities to students for self-evaluation and self-reflection. They also felt that grades assisted parents in their roles of providing guidance and making decisions for their children. The second theme was the need for teachers to consider the individual needs of a student when assigning grades. Factors of effort, participation, ability, and other student-specific situations were considered.

At the higher grades the concept of fairness was a recurring theme. Grading systems needed to be fair; it was important for teachers to consider individual needs, especially for low-ability students; it was important to formulate clear grading policies to inform students properly of the basis on which they would be graded; and it was important to use a variety of assignments on which to base grades. The results also indicated that teachers believed that grading was subjective and that the belief that grading could be objective was a delusion. For teachers of special-education students, a high degree of individualization of grading practices was reported.

Manke and Loyd (1991) concluded that teachers attempted to evaluate and report students' performance in individualized terms and that effort, improvement, and behavior were considered as they graded achievement. They suggested that these other student qualities, in addition to academics, should be measured and included in report-card grades. Furthermore, they believed that these qualities could be measured as consistently as, for example, students' processing of writing skills, thereby raising

implications for change in the way preservice teachers were trained. It was not clear whether Manke and Loyd intended that achievement and other student qualities form a single grade or multiple grades. It appears that a single grade was intended by their statement, "There is no reason why socialization to the norms of school behavior should not be measured and included along with academic performance in the report card grade" (Manke & Loyd, 1991, p. 12).

Griswold and Griswold (1992), Brookhart (1992, 1993), and Barnes (1985) focused on teachers' thought processes during the grading of students. In their study, Griswold and Griswold (1992) examined the nature of school achievement in light of teachers' expectations, beliefs, and attitudes in an attempt to clarify the nature of the knowledge base necessary for beginning teachers. In particular, they were interested in blending in the principles of educational measurement with educational psychology in a more meaningful way.

Their sample consisted of 326 subjects, of whom 78% were preservice teachers and 22% were inservice teachers. The teachers were from both rural and suburban parts of a midwestern state, and were predominately female (77%) and specialized in secondary education (82%). Two simulation stimuli were shown to the teachers. One simulation portrayed a low-ability—high-effort student and the second portrayed a high-ability—low-effort student, both of whom achieved well on a final examination. The subjects were to first decide on a grade for each student based on the four factors of achievement, effort, attitude, and aptitude, each of which was rated on a scale from 1-10. They were then asked to provide the reasons for their choices.

Statistical analyses were used to detect relationships among the grades awarded, the characteristics of the subjects, and the ratings of the importance of each grading ingredient in order to uncover patterns of beliefs. The results reported here are restricted to those relevant to the present study.

The findings suggested that effort was a strong influence in the assignment of grades. For example, in the low-ability—high-effort scenario 60% of the respondents chose to give a higher mark; in the high-ability—low-effort scenario 79% chose to give a lower mark. A greater proportion of elementary and preservice teachers awarded higher marks for the low-ability—high-effort student; however, equal numbers of elementary and secondary teachers awarded a lower grade for the high-ability—low-effort student. Teachers tended to grade harsher than student teachers did in both situations.

Overall, justification of grades given by respondents involved considerations of effort, attitude, and potential to learn. For example, when teachers gave a lower grade, they said that the student was not meeting his/her potential or that poor academic habits should not be rewarded. When they gave a higher mark they acknowledged that effort or a positive attitude made the difference. There was often a show of sensitivity, concern, or the benefit of the doubt.

Griswold and Griswold (1992) concluded that attitude, effort, and improvement were worthy of consideration in assigning grades. The study showed that teachers did not intend that grades reflect only achievement, but that they also reflect effort and attitude. They suggested the use of more than one grading indicator to separately reflect social competencies from academic behaviors, for "teachers may be reminding us that school success includes more than achievement" (p. 26). Recommendations included training teachers in ways of judging and reporting the broader social competency outcomes of schooling.

Brookhart (1992) investigated (a) the meaning which grades had for teachers, (b) the kinds of ethical reasoning in which teachers engaged when assigning grades, and (c) the extent to which meaning and ethical considerations varied with teachers who had or did not have instruction in educational measurement.

The subjects were 84 certified, employed teachers registered in graduate programs at Duquesne University. Forty teachers had completed a course in measurement instruction, whereas the remaining 44 had not. The teachers represented all grade levels, and the median number of years of teaching experience was five.

The teachers were presented with seven scenarios about grading, three of which involved students who worked to ability; two, students who missed work; and two, students who improved in scores. Accompanying each scenario was a series of choices from which the teachers were asked to select one choice which corresponded to how they would act in each situation. Then, for each option selected, they were asked to explain why they chose it.

For the three scenarios about working to ability, 81% of the teachers reported that they would base grades on the quality of work turned in regardless of effort given. However, in the case of a failing grade, 94% said that they would assign a passing grade in the case where a student's effort was evident. The scenarios involving missing work and improved scores had mixed results, but there was evidence that respondents considered social consequences, particularly in the cases of low-ability students.

Among the explanations provided for the course of action taken, the most common was that grades represented a form of payment (n=61), something to be earned for work done or effort exerted. Fewer teachers indicated that grades meant a calculated score (n=24), reflected academic achievement (n=22), or had a self-referenced meaning (n=15). The teachers justified giving particular grades by comparing students' performance in other areas, by noting their work effort, and by considering other mitigating circumstances of the class. Comments relating fairness to grade interpretations were prevalent, as were concerns about the consequences of

giving a particular grade, especially where self-esteem and self-confidence were at stake.

Though there were some differences in the meanings which teachers gave to a grade, instruction in measurement did not affect the way that teachers thought about value implications or social consequences of grades. Brookhart (1992) concluded that "teachers do not follow recommended grading practices partly because there is a conflict between the recommended practices, which concentrate on grade interpretation and meaning, and concerns with the uses of grades" (p. 5).

In a subsequent study using the same data described above, Brookhart (1993) further analyzed the meaning of fairness and of earning grades in order to provide a theory about the relationship between grading and classroom management. The results suggest that grades are related to classroom management through the concept of students earning grades by doing work. Academic achievement is only one dimension of earning a grade; following rules and trying hard are other factors of daily classroom life. In her opinion this finding was at the root of the problem of the confounding of effort and achievement in grading practices.

Barnes (1985) asked 20 student teachers and their co-operating teachers, from urban centers in western and southwestern United States, about their opinions on grading and evaluating classroom pupils, their knowledge of evaluation concepts, and the processes they used when evaluating students. The information and data were collected in a variety of ways: journal entries, conferences, interviews, classroom observations, and batteries of tests measuring cognitive, philosophical, and psychological attributes.

The main findings were that (a) all participants stated that grading was the most difficult aspect of teaching; (b) the majority of co-operating teachers were concerned about effects of negative evaluations on student achievement, attitude, and behavior; (c) for co-operating teachers, the main purpose of evaluation was to provide parents

and students with grade information; for student teachers, evaluation served the functions of motivation for students, communication to parents, a means to classify students, and a method to assess their own effectiveness; (d) co-operating teachers' evaluations of report-card grades went beyond simple averaging of numbers to "the more difficult task of making judgments based, in part, on the informal information gathering conducted on a daily basis" (Barnes, 1985, p. 47); and (e) there were no clear criteria for the evaluation (co-operating teachers reported using both grading systems and intuitive judgments based on numerous daily grades), and conflict existed between evaluation on the basis of performance and evaluation on the basis of effort. Barnes concluded that teachers "wrestled with balancing district demands for grading with pupil needs for affirmation and guidance" (p. 48).

General Reviews About Grading

The reviews in this section are related to the practical issues that teachers consider as part of the procedures of grading. For example, issues related to criteria included in grades, types of measures used to obtain scores for grades, and the determination of grades and their distributions are addressed. In addition, studies that show a comparison of teachers' actual practices with those recommended by the measurement community are reviewed. Finally, issues related to the communication and reporting of grades, grading in daily settings, and the validity of grades are reviewed.

Much of the literature contained studies in which groups of both elementary and secondary teachers participated. However, at times the results did not clearly indicate which group was referred to. It was assumed that, unless specifically stated, the results were taken across the total sample; consequently, this review reflects that assumption. In addition, at times the results from only the elementary level are

reported here; selected results from other levels of schooling are reported if they are applicable to the present study.

Procedures of Grading

The following four studies by Nava et al. (1992), Wood et al. (1990), Allal (1988), and Stiggins and Bridgeford (1985) focused on a number of issues related to the technical or procedural aspects of teachers' practice. Nava et al.(1992) examined (a) the criteria that teachers used in grading, (b) the weights given to each criteria, (c) whether the grading criteria changed when teachers considered them in the context of a specific class, and (d) whether there was consistency between what the teachers thought should be included in grading and the weight they gave to criteria in specific grading situations. Lastly, they compared elementary and secondary teachers in terms of the grading criteria they used. The sample consisted of 371 elementary and 456 high school teachers from 18 school districts representing various regions of the United States.

The survey instrument consisted of three sections. In the first section gender, grade level taught, and teaching experience were noted. In the second section 35 grading criteria were rated using a 4-point Likert-type scale according to the degree that each one should be included in grading: (1) "definitely include," (2) "probably include," (3) "probably not include," and (4) "definitely not include." In the third section the teachers indicated the approximate percentage that they gave to each grading criteria in a particular context in which they identified their own class by subject and grade level. Again the same 35 criteria were rated using 4-point scale: (1) "none, 0%"; (2) "small, 1-10%"; (3) "moderate, 11-40%"; and (4) "large, 41% or more."

Descriptive statistics that were used to address the first two questions were reported for the total sample. The five most important criteria that teachers reported

that they 'should include' in grading were unit tests, announced quizzes, essays or term papers, effort, and semester tests (median=1.0). Of these five, all were achievement related except effort. Criteria that teachers would 'probably include' were projects outside class, homework, book reports, class participation, and class exercises (median=2.0). Criteria that they would 'probably not include' were spelling, handwriting, consideration for other students, and aggressive behavior (median=3.0). They would 'definitely not include' gender, socioeconomic status, or parental involvement in school activities (4.0). The weights that the teachers gave to specific criteria varied with their importance, but very few teachers gave a large weight to any one criterion.

Factor analyses were used to examine the underlying dimensions of grading criteria under a general and a specific classroom context. The results indicated that the same four factors appeared consistently in both general and specific grading situations. The first factor included behaviors and characteristics which were perceived by teachers to enhance or deter the learning process. The second factor included assessments of achievement and academic content. Factors three (other student behaviors and noncontent skills) and four (traits that teachers considered as external to the classroom) were somewhat weaker. Correlation coefficients, which were calculated to find relationships between degree of inclusion and percentages given to each of the 35 criteria, were all statistically significant beyond the .01 level of significance, indicating that the criteria that teachers said should be included were the criteria to which they tended to give weight.

Elementary and secondary teachers differed on certain specific grading criteria. In contrast to secondary teachers, elementary teachers were less likely to include standardized tests, announced quizzes, essays, reports, homework, assignments, and group projects, but more likely to include checklists and improvement (both within and between grading periods).

Nava et al. (1992) concluded that, across all levels, teachers used both achievement and nonachievement criteria in grading and that very few teachers weighted any one criterion heavily. They also concluded that teachers showed consistency between the criteria they included in grading and their respective weightings.

Wood et al. (1990) examined (a) how experienced school teachers distributed their end-term grades across the traditional A-B-C-D-F grading scale; (b) what student characteristics contributed to student assignment to a particular grade category and how teachers weighted student achievement, effort, and potential in their grading decisions; (c) which devices were preferred in summative evaluation; (d) what teachers' attitudes towards the use of teacher-made and districtwide tests were; and (e) whether teachers teaching different types of classes differed in their grading practices. They hypothesized that teachers of handicapped students and teachers of performance-based classes would assign higher grades to more students, that these two groups of teachers would weigh improvement more heavily in their term grades, and that elementary teachers would weigh homework and classwork more than secondary teachers. A convenience sample of 258 experienced teachers (50% elementary and 50% secondary) completed a questionnaire related to the above questions.

The results indicate that the typical (median) teacher assigned A, B, C, D, and F grades to 20%, 30%, 30%, 10%, and 5% of their students, respectively. However, there were notable differences across individual teachers in their grade assignments. For example, two teachers assigned As to 90% of their students, whereas four teachers assigned no A grades to their students. Elementary teachers assigned more As and Bs than did secondary teachers, and about half as many D or F grades. Teachers of music, art, or physical education assigned more As and Bs

(70%) than did teachers of academic classes (54%). In spite of this diversity, 68% of the teachers teaching similar classes agreed that they used similar grading systems.

Sixty-five percent of the typical (median) teacher's grade was reported to have been based on demonstrated student achievement. Completion of homework and classwork was the second most heavily weighted factor; however, elementary teachers weighted this factor only about half as heavily as secondary teachers (means of 12% and 22%, respectively). Motivation, effort, improvement, potential, and participation were weighted to a much lesser degree (1-5%). Teachers of performance classes weighted achievement of academic objectives only half as heavily (a mean of 23%) as did other teachers. Participation and behavior in class (18%), attitude (12%), and improvement (9%) were more heavily weighted by these teachers than they were by typical teachers.

In summative grading, the following percentages of teachers reported certain measures as most useful for identifying whether class objectives had been met: oral questions (41%), class seatwork (38%), selection or completion questions (28%), observation of effort/motivation (22%), and problems (21%). Homework assignments and essays were chosen by 15% to 17% of the teachers. Seventy percent did not feel that standardized tests would improve education. Though they were 'pro-test,' half of the teachers reported that grades were not necessarily more valid when based on test scores.

Over 90% of the teachers stated that report-card grades should reflect how much students have learned (including homework). Further, 85% said that grades should reflect effort as well as levels of achievement, although it was not made clear if effort and achievement should be graded separately or together in one grade. Seventy percent reported that teachers who teach similar classes tend to use similar grading systems and that teachers should set different objectives for different students. It was

not clear what was meant by 'different,' although 52% of the teachers specifically said that they set different objectives for students with differing abilities.

Allal's (1988) study focused on the evaluation practices of first- through sixth-grade elementary teachers in Geneva, Switzerland. Forty-five teachers were interviewed using a semistructured format. The three strategies dealt with were implicit references used when assigning grades for the outcome of a single assessment, procedures adopted to combine information resulting from several assessments in order to determine a summative grade, and processes used to make end-of-year (promotion or placement) decisions.

The results showed that 88% of the teachers used a combination of norm- and criterion-referenced interpretations in their grading practices. For example, a grade of 4 (equivalent to our "C" letter grade) was determined by the definition of minimal objectives attained, and the other grades were determined by the distribution of the students' results. In combining information from several assessments to determine summative grades, 76% of the teachers used three or more "big" tests (teacher-made tests of substantial length covering several content chapters and/or objectives), and 74% supplemented this information with quizzes or daily completion of tasks including worksheets, homework, and observations of student participation. Group work was used by only 16% of the teachers. As far as rules for combining the information into grades were concerned, 40% of the teachers reported computing a simple or weighted average; however, the majority of the teachers made adjustments to the quantitative core. This involved a variety of elements; for example, "assessment of effort or of perseverance, . . . unrecorded and intuitive observations of a child's attitudes and work habits, [or] global judgments regarding the adequacy of the child's skills" (Allal, 1988, p. 47).

When making decisions about promotion or placement to junior high school, only 13% of the teachers gave primary importance to students' grades for promotional

purposes, and 24% gave primary importance to grades for placement. The majority of teachers made decisions based on other factors such as psychological, emotional, and intellectual development; motivation for school work; or degree of social integration. It was noted that teachers did not have systematic means of record keeping and that decisions were based largely on intuition and informal observation, in contrast to the written regulation which stipulated that decisions were to be determined by grade-point averages.

In addition to the importance of how grades were assigned and what factors teachers considered in assigning grades was the related knowledge of what types of measures teachers used in order to derive grades. Although research on testing has tended to concentrate on the role of and attitudes toward standardized tests (Stiggins & Conklin, 1992), a few studies have focused on teacher-made assessments. In one of these, Stiggins and Bridgeford (1985) conducted a survey across teachers from Grades 2-11 to determine the role and relative importance of several types of measurement in the classroom. At the Grade 5 level approximately 50% of the teachers reported comfortable use with teacher-made objective tests, published tests, and structured-performance assessments; whereas 83% reported comfortable use with spontaneous-performance assessments. For grading purposes these teachers attached slightly more importance to teacher-made objective tests (35%) than to spontaneous-performance assessments (20%). Three quarters of the teachers expressed concern about the quality of their tests and indicated the need for test improvement. There were more concerns about judging and grading students accurately as grade levels rose, indicating the increased importance placed on grades as a measure of student success as grade level increases.

A Comparison of Teachers' and Measurement Specialists' Views on Grading

Frary et al.(1992) and Stiggins et al. (1989) examined grading practices at the secondary school level. They were particularly interested in and observed discrepancies between measurement specialists' recommended grading practices and teachers' actual understanding of these practices.

Frary et al.'s (1992) specific purpose was to determine what remediation or training in measurement was needed to correct any false assumptions that teachers of high school academic subjects may have had about the meaning of grades. They investigated teachers' beliefs and practices using six broad questions: (a) To what extent do teachers interpret test scores as representing the percentage of knowledge that a student has learned? (b) how pervasive is the practice of assigning letter grades directly on the basis of percent-correct scores? (c) to what extent do teachers appreciate the need for relatively difficult tests if the ranking function is to be served optimally? (d) to what extent do teachers believe that differences in percentage-grading scales across school districts constitute real differences in standards? (e) to what extent do teachers endorse the use of factors other than achievement in determining course grades? and (f) how do teachers determine the minimum passing score for a test?

A factor analysis of the opinions of teachers revealed that teachers believed that (a) districtwide percentage grading scales were generally desirable and effective and that a percent-correct score indicated the absolute amount of a student's knowledge, (b) tests difficult enough to maximize ranking effectiveness were undesirable and pedagogically unsound, (c) extraneous factors of effort and conduct should influence course grades, and (d) minimum passing scores should be set at a fixed percentage of correct answers.

Although it appeared that the teachers generally favored a criterion-referenced approach, the results showed that few teachers reported using a criterion-referenced

approach to testing. Furthermore, whereas 41% of the teachers believed that test scores provided only ranking information, 46% reported using percent-correct scores in "conjunction with an apparently domain-referenced interpretation" (p. 9). The results also showed that teachers reported using homework and class participation as factors in determining grades.

Answers to the remaining questions were more varied and therefore inconclusive. For example, a contradiction was evident when teachers reported the belief that districtwide percentage-grading scales were desirable but that percent-correct scores could not reflect an absolute level of student knowledge. Frary et al. (1992) questioned the teachers' practice of using their tests within a domain-referenced interpretative framework. They clearly advocated a norm-referenced approach to testing, given the complex and varied nature of secondary academic subjects and the time commitment required in producing quality domain- or criterion-referenced tests. They concluded that there was widespread disagreement between what the teachers reported as practice and what measurement specialists would recommend.

Using case-study methodology, Stiggins et al. (1989) explored the nature and technical quality of assessment and grading practices of 15 high school teachers (location not reported). Information was gathered on 34 issues of teachers' grading practices. These issues were selected for inclusion in the study because they could be addressed by the measurement community in terms of specifying recommendations or best practices. Each of these issues was then compared to teachers' actual practice, and discrepancies were noted. The study focused on three areas: basic assumptions or antecedents of grading, the actual grading practices, and the effects of grading. Discrepancies were noted in 26 of the issues. For example, teachers considered ability, motivation, and effort in the formation of grades, contrary to recommendations of measurement experts. Also discrepant with recommended

practice were methods used to aggregate assessment data, procedures of setting cut-off points for determining grades, and decisions regarding borderline cases using subjective, nonachievement data. It was concluded that teachers did not adhere to recommended grading practices and that for some issues further training for teachers was needed. However, for other issues, the authors suggested that the recommended practices may have been a matter of opinion or practically inappropriate given the realities of the common classroom.

Communication and Reporting of Grades

The reporting of student achievement and progress is closely tied to teachers' grading practices, for how teachers assign grades has a bearing on how their meaning is interpreted by others. As for the formation of grades, there are concerns. The questions addressed include how teachers, students, and parents have interpreted grades (Burton, 1983; Pilcher-Carlton & Oosterhof, 1993); how the interpretation of grades can be compromised (Friedman & Frisbie, 1993; Mehring, Parks, Walter, & Banikowski, 1991; Waltman & Frisbie, 1993); and how grades should be interpreted (Friedman & Manley, 1991).

Burton (1983), in an examination of the letter-grade system, specifically addressed the rationale for giving letter grades, how they were assigned to students, the consequences of letter grades for students, and possible alternatives to grades. He found that 52% of the elementary teachers surveyed reported that the primary reason for giving letter grades was that the district requires it. Although an additional 29% stated that the primary reason was to inform parents, Burton (1983) found that, when he interviewed these teachers, the reason could have been restated as "because parents wanted it" (p. 2). Middle and high school teachers listed the primary reason for giving letters as "informing students."

About half of the elementary teachers listed numerical scores as their primary means of assigning grades, though another 45% reported using observations and professional judgment. A third reported using student participation and enthusiasm as a 'second' consideration. Middle school (83%) and high school (85%) teachers assigned grades based on numerical scores that students received on paper-and-pencil tests. Over half of these middle and high school teachers reported using observation and professional judgment as their 'second' means for assigning grades.

Half of the elementary teachers and 75% of middle and high school teachers felt that the letter grade system exerted no pressure on students. However, 50% of elementary and middle school teachers felt pressure from parents, administrators, and students themselves when assigning grades. Over half of the elementary and three quarters of middle and high school teachers believed that the letter-grade system did not negatively influence students' motivation, although many teachers were undecided whether grades had a positive influence on students' motivation or not. About 43% of the elementary teachers and 25% of middle and high school teachers believed that the letter-grade system had a negative effect on how students felt about themselves. Most of the elementary teachers (63%) felt that letter grades were accurate indicators of student learning; middle school teachers were more evenly split on this issue. Burton (1983) reported that, whereas most of the middle and high school teachers felt that the positive effects of letter grades outweigh the negative effects, elementary teachers were split on this issue.

Whereas 61% of the elementary teachers believed that there were viable alternatives to the grade system, the most favorable being checklists (79%), written progress reports (68%), and conferences (60%), 40% of middle school and 33% of high school teachers believed that there were no workable alternatives. Only a few teachers at each of the instructional levels suggested that alternatives could be used to *supplement* rather than replace letter grades.

In addition, Burton (1983) found that students across elementary, middle, and high school levels liked getting letter grades and believed them to be informative and important, and that letter grades were "a manifestation of justice and the concept of work" (p. 6). Even though parents' explanations of what grades meant were varied, over half of them stated that the grading system used was informative, and most believed that there were no alternatives to the letter-grade system. However, over half of the parents felt that letter grades should be supplemented with more teacher comments.

Pilcher-Carlton and Oosterhof (1993) found that high school teachers developed their grading systems to protect low achievers who display effort. These teachers did not fail students in such cases but considered effort and ability. Further, these teachers were aware that students were punished for low grades and therefore made allowances for students to avoid punishment. Though there was variation in how achievement and effort should enter into the grading equation (students thought that effort should be used as an incentive and not as a penalty, whereas teachers felt that effort should be used in both situations), parents understood grades to represent only achievement and felt that grades were not clear if they included other things. When probed, however, it was clear that parents of low-ability children were in favor of including effort.

Mehring et al. (1991) found that even though elementary school teachers used the same achievement and effort symbols on report cards, there was inconsistency in their interpretations of those symbols. For example, teachers assigned grades in reading based on advances made in a basal reading series (student progress) but assigned grades in other subjects based on the completion of assigned work using percentage cut-offs. A similar finding was reported by Waltman and Frisbie (1993), in that significant numbers of parents and teachers believed that grades could be interpreted according to both norm-referenced and criterion-referenced standards; in

the researchers' opinion, "two logically-incompatible meanings" (p. 17). They concluded that "there was a significant amount of variability among parents and an intolerable level of inconsistency between teacher and parents in the way grades from a given classroom are interpreted" (p. 17). Friedman and Frisbie (1993) also concluded that "report cards have serious limitations which compromise the validity of the information they convey" (p. 28) and included a set of recommendations and guidelines to follow for a more consistent interpretation of report cards.

In addition to the previously stated belief that teachers believed achievement should not be the only factor considered in grading, Friedman and Manley (1991) discovered that high school principals, counselors, students, and parents thought that teachers should consider student motivation, effort, and attitude and that these factors should be included in the final grade.

Grading of Daily Work

Though formative in nature, teachers' judgments in the daily grading of students' work was viewed as relevant to the present study. Mead (1992) examined the criteria that teachers described when assigning grades to individual pieces of mathematics work. The wide variation in grades assigned by the teachers to the same piece of work and the consistency within each teacher to retain the same grade over a period of time led the researchers to conclude that elementary and high school teachers based their grades on a particular philosophy or importance which they held about a particular grade.

Mead (1992) found that both elementary and secondary teachers considered a student's self-image and level of understanding and expressed that as a grade. For example, elementary teachers considered partially correct work. They were reluctant to give a final grade, citing the effect of grades on student motivation, and preferred instead to allow students the time and practice to improve or to give a qualitative

response to guide them prior to assigning a final grade. It was concluded that elementary teachers favored some sort of grading system which allowed students to be encouraged to continue with a desired behavior and that "student motivation is part of that concern for future student performance" (p. 18).

Secondary teachers were more likely to view student performance as "representing past or present accomplishment that needs rewarding" (Mead, 1992, p. 18). They viewed grading as necessary to form part of a permanent record for students and thought that students bore the responsibility for their own failure. They were more inclined to reward effort and improvement than their elementary counterparts were.

Whitmer (1983) described the judgment processes of five elementary teachers during the marking of their students' work over a school year. Task completion and a given standard of mastery emerged as the main judgment cue of teachers during the marking process. However, completion of work often carried a heavier weight than quality, especially for the low achiever. Whitmer also noted that low-ability students were not differentiated to as great a degree as were able students. She suggested that the demands of the classroom environment and the future placement and success of students most strongly influenced teachers' judgments and decisions. She further noted that society used marks "as measures of academic achievement against an absolute standard (mastery)" (p. 15) and that there was a discrepancy between the functions taken into account by teachers when judging students and the functions ascribed to grades by society in general.

Validity of Grades

In addition to understanding how teachers think about grades or how they actually work through the grading process, it is worthwhile to investigate the extent to which grades represent valid and reliable descriptions of student achievement. The inherent subjectivity of teachers' judgments has raised the question of whether teachers are able to form valid grades. The following three studies (Austin & McCann, 1992; Hoge & Coladarci, 1989; Leiter, 1983) are related to this topic.

In response to state leaders' concerns about grading policies and procedures in schools, Austin and McCann (1992) conducted a study in which they provided state leaders with descriptions of current grading policies and procedures. Using documents that were submitted by 144 high school districts in the state serving the focus of the study, they analyzed those policies and procedures that affected the grading practices of high school teachers of English and mathematics.

The results of the study provided insight into the extent to which a school community had a shared understanding of what grades represent. The results were given at four levels of school organization—board, district, school, and department—where appropriate. Of seven topics extracted through a content analysis, five are most appropriate for discussion here: the purposes of grades, the audiences for grades, the criteria considered in calculating grades, grading-related practices, and staff development.

At the board level the four most common written purposes for grades were to provide information (a) about student progress (82%), (b) for instructional planning (44%), (c) about a student's current level of achievement and/or performance (25%), and (d) for decision making (e.g., placement in level of course and promotion [no percentage stated]). At the district and school levels, providing information about student progress and for instructional planning was cited less frequently (40% and 38%; 20% and 10%, respectively) as a purpose for grades. However, providing

information about student achievement was cited more often at the district level (45%) and at the school level (65%) than at the board level. There were a few documents (no percentage stated) describing grades as "motivators" and "rewards."

At each of the organizational levels, the documents revealed that the most common audiences for grades were parents (77%-85%), students (54%-76%), and teachers and other school-level decision makers (10%-25%). Of the criteria recommended to teachers as the basis for calculating grades, five were identified and their emphases recorded: student performance—based on tests, assignments, homework (85%-97%); class participation—based on work habits, work completed, preparedness for class (30%-38%); and student behavior—based on attendance, attitude, and discipline (3%-23%). Only 17% of the districts established student performance as the sole criterion to be used in determining a grade; in the remaining districts, teachers were asked to use multiple criteria. For example, one board policy indicated that a final grade was determined by weighting two criteria: comprehension and skill, which accounted for 50 to 75% of the final grade; and initiative, which accounted for 50 to 25% of the final grade. Comprehension and skill were assessed using tests, reports, projects, and written and oral reports; initiative was assessed using homework, logs, and classroom participation. Only one of the 144 districts studied suggested a two-grade system, with an academic achievement grade and an individual development grade. And only one suggested that teachers consider using I.Q. in the determination of grades (Austin & McCann, 1992, p. 7).

Where districts provided documents from all levels of organization (49%), grading policies were compared across levels for consistency. Austin and McCann (1992) found that in 65% of these districts' grading policies differed across organizational levels. For example, "the documents from different levels described different criteria, different numbers of criteria, and/or gave different emphasis to selected criteria" (p. 8).

Of the districts that provided information on symbol systems (n=112), 78% reported using simple letter grades (A-F) or a combination of letters and pluses and/or minuses; 23% used numerical systems, either described in bands (e.g., 90 to 100: excellent) or without bands. Of 64 districts which submitted information on staff development, it appeared that "grading policies and practices were most often treated as an informational topic during meetings . . . [instead of] in-depth discussion of the topic or [an indication] that there was 'training' to increase the consistency with which school staff determined grades" (Austin & McMann, 1992, p. 8).

Of significance in the above study is the reported variation across the 144 districts regarding the content of their grading policies, the inconsistency with respect to the recommendation of criteria within the levels of a given district, and the practice of asking teachers to apply multicriteria to determine their grades, a practice which influences the meaningfulness of grades. In addition, none of the 144 districts provided information about helping teachers to be consistent in their grading practices.

Hoge and Coladarci (1989) reviewed 16 studies which were focused on the relationship between teachers' judgments of their students' academic performance and the students' actual performance on an achievement criterion. The judgments that teachers were asked to make varied in degree of specificity. For example, teachers' ratings of students' academic ability were considered to have a low degree of specificity, whereas teachers' judgments of the number of correct responses that students should make were considered to have a high degree of specificity. The results reported in the studies indicated generally high levels of agreement between the judgmental measures (whether they were of a low or high specificity) and the standardized achievement test scores. The authors concluded that teachers' judgments of their students' academic achievement were valid. They further noted that a median correlation (.62) "exceed[ed] the convergent and concurrent validity coefficients

normally reported for psychological tests" (p. 308). However, it is not known which grade levels were included in the studies reviewed by Hoge and Coladarci (1989) or whether teachers of elementary and secondary students were equally represented in these studies.

In contrast, Leiter and Brown (1983), in investigating teachers' grading of second- and third-graders, found that reading and mathematics achievement measured by standardized year-end tests were only weakly related to the grades awarded by teachers. Further, Leiter observed that the grades that students received in the first grade shaped the grades in the following year. Grades were also strongly influenced by student conformity to the teacher's preferred attitude and behavior patterns. Leiter concluded that the results, though equivocal, pointed to "enough evidence of nonmeritocratic inputs into grading to raise serious questions about the meritocratic pretensions of elementary school grading" (p. 18).

Canadian Studies

In contrast to the studies reviewed in the previous sections and which were conducted in the United States and Europe, grades and grading have not been examined as a distinct variable in Canadian studies. Instead, there has been a more general focus on what Stiggins (1985) termed *classroom assessment*. What follows is a review of four studies: Thiessen and Moorhead (1985), Bateson (1990), Schulz (1993), and Anderson and Bachor (1993).

Working with elementary teachers, Thiessen and Moorhead (1985) asked teachers what they thought and valued about students' evaluation. Although they focused on issues and concerns along a broader concept of evaluation than in the present study, they found that the way that teachers thought about evaluation depended on whether they had a "responsive" or an "interactive" approach to their classes. For example, when identifying the basis on which judgments in evaluation

were made, a teacher with an interactive orientation would use the experiences and progress of her students as the basis for evaluation, referred to as "accountability from inside" (p. 18); in contrast, a teacher with a responsive orientation would apply age, grade, or program norms as the basis for evaluation, referred to as "accountability to outside" (p. 18).

Bateson (1990) surveyed Grade 4 (n=321), Grade 7 (n=328), and Grade 10 (n=324) British Columbia science teachers on aspects of teaching practices, among them measurement and evaluation practices. In a questionnaire teachers were asked about the kinds of student characteristics that should be and were measured, what methods were used to collect this information, and to what extent external sources of testing were important.

All teachers in Grades 4 and 7 were asked a question regarding the degree of emphasis they placed on certain measurement methods for deciding on a final evaluation for their students. The results showed that Grade 4 (48%) and Grade 7 (54%) teachers gave much emphasis to teacher-made objective tests. They also gave a fairly heavy emphasis to projects (16% and 18%, respectively) and to experiment write-ups (16% and 27%, respectively). In addition to the difference in emphasis given to experiment write-ups between Grades 4 and 7 teachers, Grade 4 teachers gave more emphasis to anecdotal records (11%) and oral tests (5%) than did Grade 7 teachers (6% and 2%, respectively). Only 1% of the teachers at each of the two grade levels placed much emphasis on the provincial achievement test; 62% of Grade 4 and 58% of Grade 7 teachers placed no emphasis on the provincial achievement test. In a similar question, the Grade 10 teachers were asked to rate the importance of certain methods of measurement for a final evaluation; however, the results were similar to those reported for the Grades 4 and 7 teachers in that the Grade 10 teachers gave the most importance to teacher-made objective tests and little reliance on the provincial achievement test. In addition, attendance and class

behavior became more important for a final evaluation at the secondary level than at the elementary level.

Further, Bateson (1990) found that although the curriculum stated that the four goals of attitudes, skills and processes, knowledge, and critical thinking received equal emphasis in the junior-secondary program, when measuring and evaluating, the teachers emphasized the knowledge component; they paid very little attention to the affective domain. Bateson concluded that teachers had neither the skill nor the confidence in techniques for assessing attitude and critical-thinking skills.

Although there was no report about what role school- or districtwide exams played in the final grade of Grade 4 or Grade 7 students, when Grade 10 teachers were asked whether other measurements played a part in determining a student's final letter grade, two thirds of them indicated that they used a schoolwide and/or a districtwide exam. One third reported that they had the sole responsibility for a final evaluation. One third of these teachers reported using preset distributions of marks. Although 13% of the teachers had no expectations of marks for any class, the remaining teachers expected approximately 8% of the students to get an A, 10% to get a B, 53% to get a C or a C+, 20% to get a pass, and 8% to fail. Bateson (1990) described these distributions as very similar to those expected on a provincial examination in Grade 12 English. He further noted that the "lack of use of available, curriculum-referenced, quality objective tests is a finding which should cause some concern" (p. 50).

Schulz (1993) conducted a study to determine the grading and reporting policies and practices used by schools and to determine how grades and other forms of evaluation were communicated to students, parents, and others. Data were collected from 22 superintendents across Newfoundland and Labrador by questionnaire and supplemented with interviews. Superintendent results were obtained for four levels: kindergarten, primary (Grades 1-3), elementary (Grades 4-6), and intermediate

(Grades 7-9). Only the results for elementary students in Grades 4-6 are reported here.

The superintendents were asked to rate each of 13 student characteristics (generally divided into cognitive and affective based) to be used in determining a grade on a scale from 1 (not at all important) to 4 (essential). They rated the four listed cognitive characteristics—achievement (3.4), process skills (3.7), critical-thinking skills (3.4), and communication skills (3.7)—high. In addition, two of the affective characteristics—social development (3.6) and ability to work independently (3.2)—were rated high. Effort, attitude, punctuality, care and neatness, and classroom behavior and co-operation were rated moderately high (2.7-2.9).

In a second question the superintendents were asked to rate the importance of assessing and reporting various aspects of student learning. The results indicated that all the stated subject areas (2.9-4.0) and aspects of human development (3.5-3.8) were important to assessment and reporting. Interestingly, oral and communication skills (3.9), research and process skills (3.8), and social-development skills (3.7) rated higher than did French, health, and physical education (3.1 each).

The most favorable form of reporting aspects of student learning was the narrative or anecdotal format (62%-80%). The second most favorable was some form of grading format; for the reporting of achievement in subject areas the letter grade was the most popular (47%-60%), followed by checklists (45%-57%) and percentages (34%-38%). For other areas of cognitive and affective learning, the results were checklists (20%-40%), letter grades (10%-26%), and percentages (2%-33%).

Anderson and Bachor (1993) conducted a study to investigate the assessment experiences and practices of stakeholders (teachers, parents, students, and administrators) at the elementary school level in British Columbia. A stratified sample of 10 school districts with two schools from each district participated in the study. Within each school two classes at each of the Grades 3/4 and 6/7 levels were

used. Focus-group interviews were conducted separately with groups of students, parents, teachers, and principals on four themes: information collecting, collation and storage of information, analysis and interpretation, and reporting. The results were reported for each group separately, and only summaries were available. The main results from the teacher, parent, and administrative groups are reported here.

Teachers used observation as the primary assessment method, with work samples, tests, and student self-evaluation forming other common assessment practices. Assessment was not discrete but a 'running record' of what happened in the classroom; it tended to be "rather intuitive with the teacher knowing tacitly what the program is about and the general developmental trends of children at this particular age, and locating the students within it" (Anderson & Bachor, 1993, p. 3). Tests and the aggregation of scores to produce averages were more common at the Grades 6/7 level than at the 3/4 level; however, scores on teacher-made and standardized tests were reported to parents at conferences at the discretion of the student or teacher.

The main reasons for assessing students were to monitor a student's progress in terms of curriculum and to identify a student's position in relation to the goals of schooling. Progress was interpreted in relation to these goals, in relation to past performance, and to a lesser degree in relation to others. Teachers generally rejected letter grades and expressed an aversion to comparison of students, citing negative effects, and preferred the narrative report even though it was not well received by all parents or all students. Some reports provided descriptive accounts of student progress with no evaluative comment. However, some teachers saw value in comparing or using grades as a benchmark.

According to Anderson and Bachor (1993), parents viewed assessment largely from the perspective of reporting. They expressed concerns about the practice of using anecdotal reports. Parents wished to see "succinct evaluative reports on the

achievement of their children in the academic areas of reading, writing, mathematics" (p. 12). They wanted to know what was expected of their children and saw comparisons as useful and informative, especially as their children grew older. Other parents who viewed narrative evaluation positively stated that the method used to evaluate their children was not as important as the teacher who was in charge of evaluating. They were satisfied if the teacher's comments and descriptions were specific to their child and were "balanced with information on the academic status of the child" (p. 13).

The administrators were highly supportive of the focus on classroom assessment: student self-evaluation, reduction of standardized testing, and narrative reporting. Generally, the only district policy regarding classroom assessment involved the frequency and format of reports to parents. Most schools did not have a specific policy with regard to classroom assessment; the principals provided guidelines which were developed in collaboration with teachers. In-services for teachers with regard to development and implementation of assessment initiatives were co-ordinated at the school level.

Summary

Taking the literature as a whole, the following list summarizes the major points:

1. In the few studies that related specifically to teachers' beliefs about their grading practices, the topics addressed the meaning that teachers had of grades and the consideration of other student characteristics, particularly effort, that teachers felt should be included in grades. Factors related to students' ability, social consequences, and judgments made on an individualized student basis were common themes.

2. Studies related to the practical issues of grading revealed that teachers (a) used both achievement and nonachievement criteria when formulating grades,

(b) preferred and used a variety of teacher-made measures in determining student grades, and (c) employed different patterns of grade distributions for their students.

3. Studies showed that teachers' grading practices did not always conform to those recommended by measurement specialists.

4. Studies revealed that grades were interpreted differently across teachers and parents. Although letter grades remained the most popular, the results concerning their effectiveness, validity, and continued use were mixed.

5. Grading policies either differed across administrative levels of schooling or were generally implemented in flexible or nonspecific ways.

6. Overall, several studies concluded that a teacher's personal philosophy strongly influenced teachers' grading practices.

Chapter III

Research Design

This chapter first describes the subjects and the selection of the sample, followed by the development and testing of the research instrument. Following this, a description of the procedures for the main research, which included data collection, and coding and analysis of the teachers' protocols, is presented.

Subjects

The population in this study was Grades 4-6 teachers in a large urban school system in Canada. Following the granting of approval by the Research and Ethics Committee of the Department of Educational Psychology at the University of Alberta (see Appendix A), an initial sample of 120 teachers was randomly selected by staff located in the central office of the district. Those selected were each given notification by two letters which were circulated together through the district's internal mail-delivery service in late March of 1994 (see Appendix A). The first letter was written by the researcher and contained the nature of the study and a request for each teacher to contact her to participate in the study. The second letter, from the district's central office, indicated to those selected that the researcher had permission to conduct the study in the district. A subject's identity was revealed to the researcher only when that teacher contacted the researcher by telephone to volunteer to participate in the study. The teachers who called were assured of anonymity during the length of the study, as well as given the option to discontinue without prejudice.

Though requested that the sample include only Division 2 teachers, this service was not provided due to the inability to access and retrieve the information directly from the system's main computer file. Consequently, the sample included Division 1

teachers, administrators, librarians, special-education teachers, counsellors, and persons on sick leave, in addition to Division 2 teachers.

A total of 33 (27.5%) teachers contacted the researcher by phone. Fourteen were found to be ineligible and consequently were not interviewed. Of the 14, 3 were teachers of Grades 1-3, and 2 reported that they used portfolio assessments rather than grades as their methods of evaluation. The remaining 9 declined to participate either during initial contact or prior to being interviewed. Of the remaining 19 teachers, 13 were Division 2 teachers, 2 were special-education teachers, 1 was a librarian, 1 was on sick leave, and 2 were Division 1 teachers. Because all of the non-Division 2 teachers had previous experience at this level, it was decided to allow them to proceed with the study. None of the participants requested leave of the study, once initiated.

Because Division 2 teachers comprised approximately 38% of the total number of teachers in the district, it was estimated that 45 teachers of the total initial sample of 120 were currently teaching Grade 4, 5, or 6 students. Consequently, the response rate of Grades 4-6 teachers was estimated to be about 31%.

With two exceptions, all teachers had more than 20 years of teaching experience. Two teachers had each taught for 14 years. The participants included 10 females and 9 males. Though they came from schools in various geographic and socioeconomic areas within the district, there was a concentration of responses from the lower socioeconomic areas of the city. Four teachers had taken between one and three measurement courses, and the remainder had no formal training in measurement.

Research Instrument

Development of Interview Schedule

A semi-structured interview was developed to obtain the teachers' beliefs about grades and their grading practices. Development began in January of 1994 and was completed in three months.

Sources. The interview consisted of nine broad questions which dealt with issues identified from the literature, measurement theory, and practical teaching experience. The first question, in which teachers were asked to provide their meaning of a grade or the process of grading, and the eighth question, in which they were asked about difficulties they encountered during grading, were formulated from the work of Manke and Loyd (1991). Questions 4a and 4b, which asked teachers to identify student characteristics that should be included in grades and how they should be reported, respectively, and Questions 5b and 5c, which asked teachers their methods of combining scores and determining cut-off points, respectively, were derived from the research of Stiggins et al. (1989) and adapted to the elementary setting. Question 6a, which asked teachers whether ability should be taken into consideration in grading, was derived from Griswold and Griswold (1992). Question 9, which asked teachers to choose measures from which achievement grades could be derived, was adapted from the work of Nava et al. (1992).

Ideas for the following questions were derived from measurement theory: Which audiences should grades serve, and what information should they contain? (Question 2), Which interpretative frameworks are most appropriate for grading? (Question 5a), Which symbols are most useful for reflecting student performance? (Question 5d), and Should grading practices be consistent across programs and subject areas? (Questions 6b and 6c).

Ideas for the remainder of the questions or subquestions were derived from the researcher's personal experience. For example, Questions 3, which addressed

possible influences and concerns of grades, is a recurring issue in practical teaching. The consistency of grading practices across schools and districts (Questions 6d and 6e) is a current issue in the researcher's district.

Description of questions. Questions 1, 2a, 4d, and 8b were open-ended and designed to capture free responses; all other questions were structured and directed to specific issues. With the exception of Questions 1, 7, and 9, all others contained subquestions that probed an issue more deeply for fuller understanding. Questions 3a, 3b, 3d, 5b, 6a, 6b, 6c, 6d, 7, and 8c were designed to obtain a "yes" or "no" response and reasons of choice. In Questions 3e, 5a, and 5c teachers were asked to select an option and give reasons for choosing a particular option. Questions 2b, 4a, 5d, and 9 were designed to obtain written responses. In Questions 3c, 4b, 4c, 5e, 6e, and 8a teachers were asked to supply additional information or give opinions. Prompts were of two types: written and oral. For example, when responses were to be recorded on paper, teachers were given appropriate instructions and asked to read a card showing the question and options from which to choose. Prompts were also given when teachers needed further explanation or when it was believed that questions were misinterpreted. The format and the questions comprising the instrument are located in Appendix B.

Review of interview schedule. A total of six persons reviewed the interview schedule and suggested changes or corrections. Four of these were colleagues in the teaching profession, two principals and two classroom teachers. The fifth person was a graduate student in educational psychology, and the sixth was a professor of educational measurement. Each reviewer was provided a copy of the interview schedule and asked to identify areas which were not clear or not covered. A discussion with each reviewer was held. Based on the full set of information, the interview schedule was altered to make it more clear.

Field test. Due to time constraints, an informal field test was undertaken just prior to conducting the interviews. A sample of three teaching colleagues were interviewed and tape-recorded. During and after each test the researcher made notes about the proceedings. For example, pace and tone of voice were noted. These three teachers judged the appropriateness of the questions and tasks, the length of the interview, and the manner in which it was conducted. Questions on which there was some ambiguity were altered using the comments of the three teachers.

Procedure

Data Collection

All interviews were conducted at locations convenient to and chosen by the participants. Sixteen teachers were interviewed at their respective schools following class dismissal, and three were interviewed in their homes. Each interview was completed in one sitting of approximately one-and-one-half hours and was audiotaped for later analysis. All interviews were conducted between April 11 and May 17, 1994.

Each interview began with the teacher reading the study description, after which he/she was asked to sign the consent form (see Appendix C). Each teacher was asked the sequence of questions in the same manner; and prompts, as described earlier, were given as necessary. Following the first question, in which teachers were asked to state in their own words what they perceived a grade or the process of grading to mean, the researcher provided a focus for the study. A card showing the illustration in Figure 1 was given to teachers. The purpose of this step was to show the teachers that grading is one particular aspect of evaluation with specific features.

The Evaluation Continuum

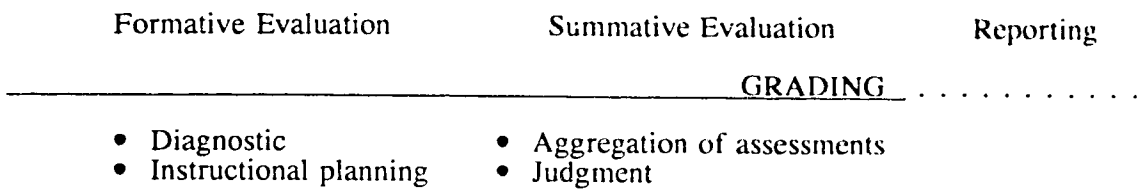

Figure 1. The position of grading along the evaluation continuum.

It was explained to the teachers that grading often connotates a broad evaluative interpretation. However, the teachers were asked to consider a more narrow definition, as indicated on the evaluation continuum (see Figure 1). A grade is synonymous with the final judgment that occurs as part of summative evaluation. Consequently, the teachers were asked to disregard those aspects of evaluation associated with formative evaluation.

Once the teachers had completed their responses to the first question, they were provided with a copy of the working definition of a grade and grading, adopted for this study. This copy, which appeared on a 5" x 8" index card, was kept in full view for further reference, should it be required. They were asked to read it carefully in order to establish a context for the questions that followed. Teachers were also provided with the following definitions or terms as they were required: *achievement* and *nonachievement* (Question 4b); *preset distribution, preset standard,* and *empirically derived* (Question 5c); and *growth* (Question 5e). A card showing a hypothetical set of class scores was also provided to teachers for answering Question 5c. As indicated earlier, some questions required written answers. Options to Questions 2b, 4a, 5d, and 9 were written on similar cards and shown to the teachers. After the teachers read the options and reflected on the responses that they wished to provide, the researcher recorded the responses. All instructions given to

the teachers pertaining to the discussion of the above questions are shown in brackets in the appropriate locations on the interview schedule (see Appendix B).

Coding and Analysis

Each audiotape was replayed immediately following the interview in order to transcribe the protocols onto individual interview-schedule sheets. Notes containing pauses, difficulties, or other pertinent information about the manner in which the teachers answered were also made. Coding and analysis began midway through data collection. The model outlined by Miles and Huberman (1984) provided a guide for the qualitative analysis. The three components of data reduction, data display, and conclusion drawing/verification formed an interactive process during the analysis.

Data reduction included transforming the raw data by the process of "doing summaries, coding, teasing out themes, making clusters, making partitions, writing memos" (Miles & Huberman, 1984, p. 21). The steps followed in reducing data are illustrated in Figure 2.

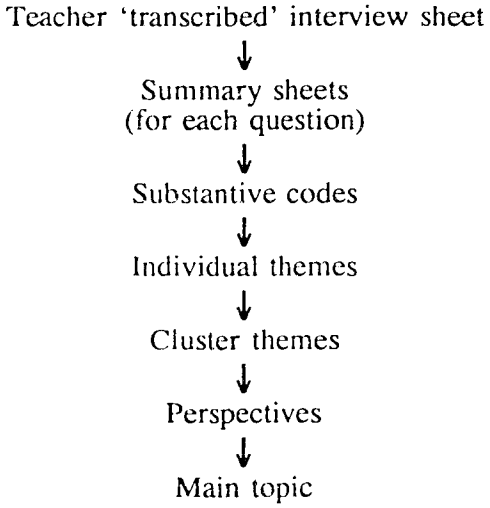

Figure 2. Steps showing sequence during data analysis.

Each teacher's transcribed interview-schedule sheet was first read in order to 'pull' usable narrative information from each question. For each question or subquestion a summary sheet was drawn up containing the relevant responses that each teacher made to a given question or subquestion. An example of a partial summary sheet for Question 1 is given below. As indicated, each teacher was given a P number, and all subsequent responses of a particular teacher were identified by that number for the duration of the study.

> P1: A level of performance. . . . When assigning a grade to a student, we say a student is capable of performing at a certain level.
>
> P2: Evaluation of how well they are doing. . . . They have an idea of my expectation.
>
> P3: A level of achievement as defined by the Dept. of Education, . . . what is expected of a child at a particular point of time.
>
> P4: A curriculum-specific material, . . . a mark that has a meaning based on these things I tell you about. . . . If you reach your objective, it's worth [this percentage for the report card].

As these summary sheets were compiled they formed the first step of data reduction and provided codable data. For each question or subquestion the responses were categorized through the process of substantive and open coding. *Substantive coding,* based directly on the data, refers to summarizing the thoughts or actions inherent in any given response. The code is a descriptive label that best defines or 'names' the concept it is describing (Miles & Huberman, 1984). For example, responses to Question 1 from five of the teachers—"to others in the class," "to a group," "to other students," "mark on a curve," and "place people in categories"—indicated that these teachers perceived grading in a norm-referenced framework. The descriptive label *relation to norm* was affixed to this particular concept.

Open coding refers to the categorization of a response into as many substantive codes as possible to ensure that concepts are not missed (Strauss, 1987). For example, the response by one teacher to Question 1, "[Grading means] evaluation and

communication to students and parents [of] how a student is doing compared to earlier," suggested three concepts. Consequently, three codes were used to capture the essence of the teacher's thoughts: judgment process, communication tool, and relation to self. A complete example of substantive codes obtained by the process of open coding of the summary sheet for Question 1 is shown in Table 1 and forms the second stage of data reduction. As indicated in the table, six substantive codes were derived from the responses of the 19 teachers. The responses used to define each substantive code are listed below that code. To facilitate later analysis, the responses were identified by each of the respondents.

Constas' (1992) two-dimensional model of category generation, designed to enhance the credibility of qualitative analysis, provided a further guide during category development. Particular attention was given to the origination, verification, and nomination of the categories (first dimension) and the temporal designation—before, during, or after the data collection—of the categorization process (second dimension). In the present study the categories originated from the narrative text of the participants; that is, after the data were collected. The researcher created each category based on some consistent function that was reflected in the responses. As previously indicated, it was given a name describing this function or concept.

As shown in Figure 2, substantive codes that emerged from the data were grouped to form individual themes at the third stage of data reduction. A theme indicated a pattern of substantive codes that were similar in meaning or had a strong relation to one another. For example, the three substantive codes *relation to criterion, relation to norm,* and *relation to self* were combined into one individual theme, *performance referent,* to reflect teachers' beliefs that grading involves a comparative framework. Substantive codes that could not be clustered with other substantive codes formed individual themes on their own. For example, the substantive codes *judgment process, communication tool,* and *symbol* each formed an

Table 1

Substantive Codes Obtained by Open Coding of Summary Sheet

1. Judgment process		2. Relation to criterion		3. Relation to norm	
P2	Evaluation	P1	To standard	P10	To others in class
P1	Awarding	P2	To teacher standard	P11	To a group
P5	Evaluating	P3	To ABED standard	P15	To group
P8	Evaluation	P4	To objectives, curriculum-specific material	P18	Other students
P9	Evaluation			P10	Mark on curve
P11	Evaluation	P5	To material taught	P13	Place people in categories
P12	Evaluation	P11	To material covered	P6	Put them in order, put students in chronological units
P14	Evaluation	P12	To objectives		
P15	Evaluation	P16	To criteria		
P6	Evaluating	P14	To standardized tests		
P13	Identifying strengths and weaknesses	P17	To objectives, curriculum		
P18	Different evaluations	P14	To curriculum		
P18	Rating, measuring	P15	Grade against standards		
		P15	Determine mastery of specific set of skills		
		P1	A level of performance		
		P3	A level of achievement		
		P8	To a level		
		P19	This level		

4. Relation to self		5. Symbol		6. Communication tool	
P8	Comparison across time	P4	A mark	P2	Child knows
P14	To general intelligence	P9	Mark, letter grade	P4	Child aware
P18	Development over time	P10	Mark	P8	Communication to parent, student
		P13	A mark, a percent	P17	Where child at
		P16	A percent, average	P14	For parent
		P7	Mark or average	P7	Parent understands how kid doing
		P15	Percent, letter grade	P19	To guide them
		P16	End result		
		P18	A measurement		
		P19	Marks		

individual theme. To illustrate the formation of individual themes, the six codes derived from the first question discussed above formed four individual themes, as shown in Figure 3.

Clustering is a tactic applied to many levels of qualitative data as a process of moving to higher levels of abstraction (Miles & Huberman, 1984). As shown in Figure 2, individual themes were clustered and identified at the fourth stage of data reduction. A cluster theme indicates a pattern of individual themes that have a strong relationship to one another. For example, the four individual themes in Figure 3 were grouped together to show what grades or the process of grading conveyed for the teachers. The cluster themes, which represent higher conceptual levels, formed the beginning of a hierarchy of concepts, shown in Figure 3 as a tree-like display known as a dendogram.

For questions that asked for specific information—for example, giving a 'yes' or 'no' response—the results were first divided into two groups to reflect the different viewpoints. From the comments that accompanied these views, substantive codes, individual themes, and cluster themes were derived in the same manner as described earlier for the open-ended questions.

As shown in Figure 2 and explained more fully in Chapters IV, V, and VI, two additional steps were followed. Cluster themes were at times divided into two perspectives to make the task of reporting results easier. For example, cluster themes that represented contextual factors inherent in teaching were divided into two perspectives to reflect an internal and an external environment and discussed separately, thus serving an organizational purpose. The main topic represents the highest conceptual level and was identified at the sixth and final stage of data reduction.

As indicated earlier, the three components of data analysis developed concurrently. As data were reduced, frequencies and hierarchies were displayed in

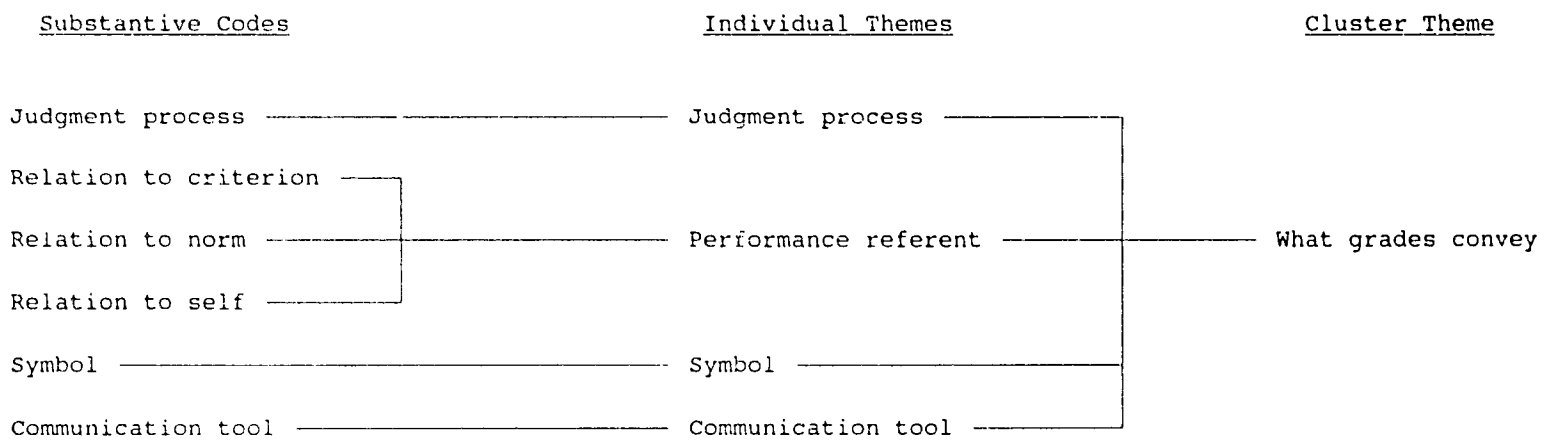

Figure 3. Formation of a hierarchy of concepts using a dendogram.

tables and dendograms. These displays contributed to the organization and presentation of the findings. The results were presented at the cluster-theme level and further discussed in terms of their component individual themes. To support the findings, responses that formed substantive codes were frequently included throughout the reported results in the form of direct quotations. Through the technique of memoing (Miles & Huberman, 1984), ideas about the substantive codes, cluster themes, and their relationships were continuously recorded as a way of conceptualizing and structuring meaning from the data. As these thoughts and patterns evolved, they formed the basis for interpreting and drawing conclusions from the data.

Preliminary Analysis

Following the processes outlined above, each of the questions or subquestions was first examined for the teachers' understanding and clarity of response. With the exceptions of four subquestions, the teachers appeared to understand the questions and provided responses that were usable. The four nonusable subquestions were: Should grades reflect the current status of student achievement, or should they be used as guides for students towards some future course of action? (3e); Do teachers, in actual practice, separate achievement and nonachievement factors when grading? (4c); How should student growth be included in the reporting of student progress? (5e); and Do you feel that guidance is needed at the school/district level for developing staff expertise and confidence in grading practices? (8c).

The systematic analysis of the responses to the remaining questions resulted in the formation of three basic topics that helped to explain the teachers' beliefs about grades and their grading practices. These topics are functions of grades, processes of grading, and contextual factors inherent in the practical teaching environment. The

results of the study, presented in Chapters IV, V, and VI, were organized around these three main topics.

Chapter IV

Teachers' Beliefs About the General Meaning of Grades

Overview of Results

As the process of data analysis neared completion, there was increasing evidence that the questions concerning the beliefs of teachers' grading practices seemed to fall naturally into three main topics: (a) the functions ascribed to grades, (b) the context in which functions are formulated and the processes are implemented, and (c) the processes used to obtain data and information from which grades are derived. As shown in Figure 4, which is a representation showing that function, context, and process are components which contribute to the meaningfulness of grades, these topics are not independent of one another. Each of these topics is first described briefly prior to reporting the findings in detail.

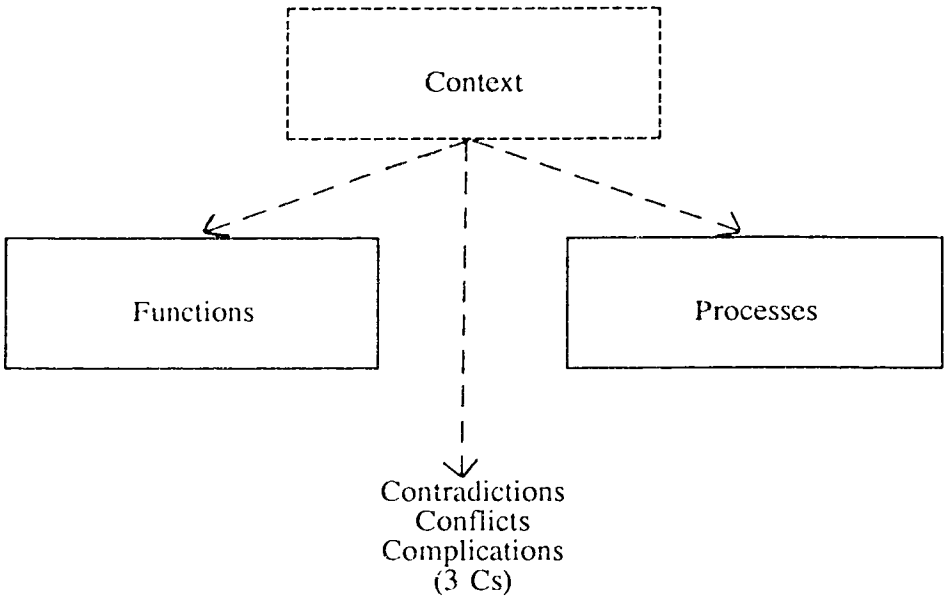

Figure 4. Components that contribute to the meaningfulness of grades.

Functions

The functions ascribed to grades by the 19 teachers interviewed reflect issues related to the communication of grades—what things are required within a particular grade that should be communicated, what symbols best reflect the meaning which teachers wish to convey and to which audiences, and what purposes grades are meant to serve for students.

Process

The process ascribed to grading by the teachers reflects the interpretation system that teachers use for assigning grades and the various procedures that they employ or think should be adopted in order to provide meaningful grades. Descriptions of how teachers choose appropriate grading frameworks, how they select measures to use for establishing grades, and how they work or would work through the procedures for determining grades are described.

Context

The ultimate goal in grading is to provide (as derived from one or more of the functions of grades or their processes) concise and accurate meanings which will be useful in matters of decision making. As shown in Figure 4, contextual factors often exist and form a mediating variable (indicated by the dotted lines) in this theoretically 'pure' two-dimensional system of grading, thereby potentially affecting the meaningfulness of grades. For example, a given process may not have served the function of a grade because of some extenuating circumstance. Due to the complexities that contextual factors presented, various contradictions, conflicts, and complications (3 Cs) arise and as such are explained by the teachers' philosophies and the environments in which teachers work.

At times it was difficult to separate the functions of grading from their processes, due to the inherent integrative nature of grading. For example, the process of assigning grades in different subject areas depended upon the symbol which best served the function of effectively communicating information to an audience. However, for clarity and convenience each of these two topics is discussed separately.

It was not possible, however, to separate contextual issues fro. . each of the other two dimensions because the overlap was clearly discernible in the subjects' responses. It can be argued that teachers, in many instances, could not or did not isolate one or two factors when assigning grades but made decisions based on the completeness of information. For example, concern of negative consequences for a child precluded a consistent, pervasive grading procedure applicable to all students. Or failure to apply consistent grading practices across subject areas, which in turn influences the meaningfulness of grades, was influenced by a particular teacher's educational philosophy that certain subjects do not lend themselves to similar grading practices. Therefore, the context which may have affected the way teachers thought about the functions of grades or their processes and the responses they gave have been incorporated into their respective chapters.

As alluded to above, contextual factors were seen from one of two perspectives: an internal environment which consists of a child's traits or characteristics and potential consequences bearing on that student as a result of these given attributes, and an external environment which consists of teachers, parents, peers, or conditions of the system which affect a child's internal environment. For ease in describing results concerning contextual issues, these two perspectives were separated.

The remainder of Chapter IV focuses upon the teachers' responses to the first question asked, which related to the general meaning of grades. The next two chapters deal with the responses to the remaining questions which, unlike the general nature of Question 1, focussed on specific grading practices. Chapter V focuses upon

the responses and interpretation of teachers' beliefs about the functions of grades and the effects of the internal contextual environment. The processes of grading as derived from the teachers' responses and comments and their external contextual influences are examined in the third chapter in this sequence.

The General Meaning of Grades

As indicated earlier, the teachers were first asked to comment on what a grade or the process of grading meant to them (Question 1). The progression of analysis leading to the identification of the general meaning of grades derived from the responses of the 19 teachers to this question is displayed in Figure 5. As stated previously in the methods chapter, substantive codes were first derived from the responses and comments of the teachers interviewed and identified at the second stage of data reduction. It will be recalled that the six substantive codes listed in Table 1 form the four individual themes and constitute the first cluster theme (see Figure 5).

As shown in Table 2, 17 teachers indicated that a "grade or the process of grading" involved describing performance in terms of one of three frameworks. Thirteen teachers referred to grading relative to a criterion, six referred to grading relative to a norm, and three referred to grading relative to oneself. In the case of criterion referenced interpretation, different terms were used to describe the criterion: "standards," "curriculum objectives," "material taught," and "set of skills," for example. For 13 teachers grading meant making a judgment. Teachers commented that when grading they were "evaluating," "rating/measuring," or "identifying strengths." Nine teachers referred to a grade as a symbol; for example "a mark," "a letter grade," or "a percentage." Seven teachers indicated that grades serve as a tool for transmitting information to an audience; in particular, to a parent or a student.

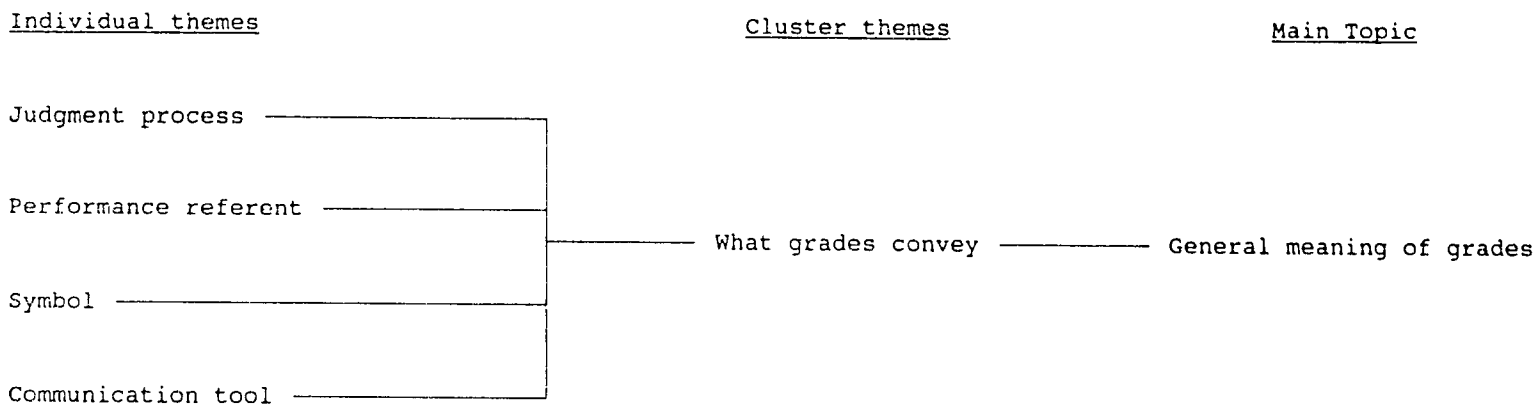

Figure 5. Progression of analysis identifying the general meaning of grades.

Table 2

Teachers' General Meaning of Grades

Theme/substantive code	n[a]	n[a]
Performance referent	17	
Criterion-referenced		13
Norm-referenced		6
Self-referenced		3
Judgment process	13	
Symbol	9	
Communication tool	7	

[a]Total possible responses = 19.

As indicated by responses to this open-ended question, teachers seemed to have a good understanding of what grades were or what they should convey to an audience. They recognized that grading involves primarily an evaluative process for the intent of communicating information using symbols and that a comparative framework for the interpretation of grades is needed.

Following the responses to Question 1, the teachers were provided a card on which was printed the researcher's definition of a grade and which they could refer to during the rest of the interview:

> A symbol or mark (i.e., number, letter, word) which represents a value judgment made by a teacher concerning the relative quality of a student's achievement of learning objectives during a specified period of instruction; a mark. Most often thought of as a percentage or a letter, a grade also includes levels—for example, at, above, or below grade level—or categories—for example, E, S, or N (excellent, satisfactory, needs improvement). A grade is a 'shorthand' language for communicating evaluative information about students. The meaningfulness of grades depends on the extent to which a school/community has a shared understanding of what they represent.

The teachers either nodded approval as they read the definition or stated that they believed in or interpreted grading in the same way. It was clear that the majority of the comments expressed by the teachers when answering Question 1 were

consistent with this definition. This was beneficial, not only because there was agreement, but also because this definition gave the teachers the basic meaning to which subsequent questions in the interview were referenced.

Chapter V

Teachers' Beliefs About the Functions of Grades

Introduction

In educational settings, where so many variables exist, seldom can teachers say that a specific practice occurs. Grading is no exception. While the majority of the teachers possess a basic philosophy and expressed, throughout the course of the interview, what appear to be solid beliefs about their grades and grading practices, in reality what exists is a compromise between these beliefs and the context in which they work. As stated earlier, contextual factors often mediate teachers' approaches to grading; and, as will be seen, they modify certain aspects of their practice.

Whereas in the previous chapter the general meaning of grades was presented as a way of providing insight into the teachers' understanding of grades, in the remaining interview questions they were asked to provide responses to questions about specific beliefs and practices. It was here that contextual factors became evident. As will be seen, the responses that some teachers made to one question disagreed or otherwise contradicted a response to another question. For example, when asked if student ability was a factor to consider when grading, three of the teachers who said "no" later responded to the question concerning the difficulty of grading students of varying abilities by indicating that there was a need to consider students of lower ability.

Presented in the next two chapters is a description of the content as gleaned from the responses of the teachers to these questions. The findings related to the functions of grades are presented in this chapter, and the findings related to the process of grading are presented in the following chapter.

Functions of Grades

The progression of analysis leading to the identification of the function of grades as one of the three main topics is displayed in Figure 6. The individual themes were derived from the analysis of substantive codes formed from the responses to various questions, as indicated earlier. These themes were then combined according to the similarity of shared ideas into "cluster" or higher order themes which, collectively, represent the main components of the main topic function of grades. For example, teachers perceived grades to provide three purposes for their students: awareness of the students' progress, a means by which students may monitor their behavior in order to receive the maximum grades possible, and encouragement. Taken together, these three ideas represent the higher-order theme of purpose/utility of grades.

As shown in Figure 6 teachers interpreted and described the function of grades in four ways: (a) what purpose(s) teachers hoped grades would serve their students, (b) to what audiences grades should be communicated, (c) in what forms grades should be communicated or reported, and (d) on what student characteristics the grades should be based.

As shown further in Figure 6, teachers accommodate contextual factors relating to the students' internal environment. The students' ability, motivation and other predispositions toward schooling, and the presence of negative consequences influence teachers' perceptions of the functions of grades. Teachers often modify their grading practices in the presence of these extenuating circumstances. Further discussion of each of these cluster themes in terms of their component individual themes follows.

Purpose/Utility of Grades for Students

When specifically asked if they hoped to influence their students with their grades (Question 3a), 16 teachers commented that they believed that grades were useful and that they hoped grades would influence their students. From their

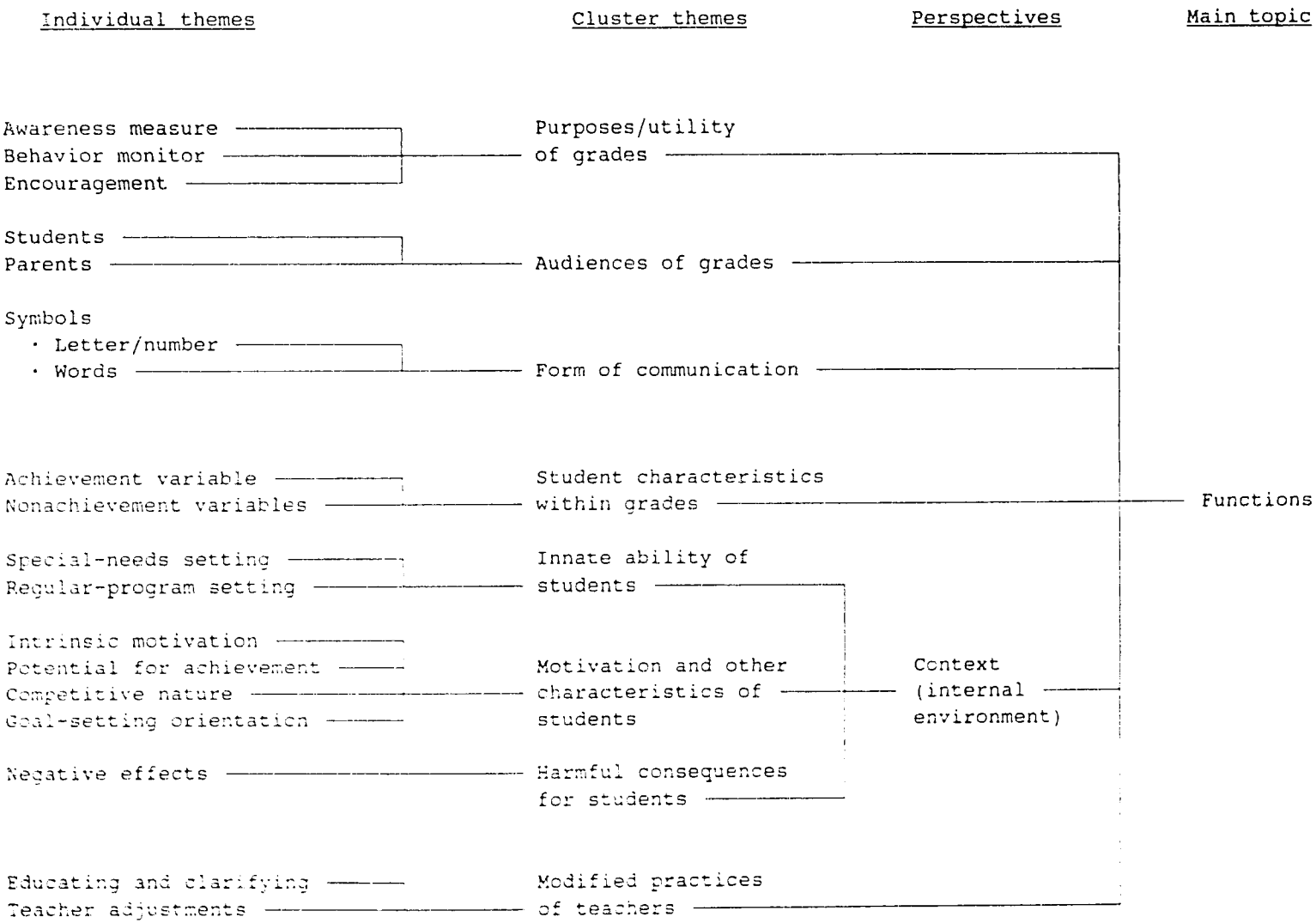

Figure 6. Progression of analysis identifying the function of grading.

responses, three individual themes which constitute the first cluster theme (see Figure 6) were identified. Twelve teachers suggested that grades serve an awareness function and provide a vehicle for monitoring student progress. One teacher remarked that a grade was "a signal, a flag [to a student]: . . . 'This is how far you've come.'" A second teacher proposed that a grade would "affirm what they know." A third suggested that grades would allow students "to see how well they met the objectives [of instruction]."

Eight teachers hoped that grades might prompt students to choose an alternate behavior should their grades be weak; for example, to "refocus," "re-evaluate [their] approach," or "choose to try harder." Lastly, eight teachers reported that they hoped that grades would serve an encouraging role. One of these teachers commented that a grade is "a reward . . . [that serves] to encourage [students] to continue to strive." Another commented that grades are useful "to cheer them on." A third said that grades encourage students "to build on [their] strengths."

Audiences for Grades

When teachers were asked to identify the primary audience or audiences to whom grades should be communicated (Question 2), 17 of the 19 teachers identified students and parents. Of these, four believed that parents were the primary audience; further, two teachers felt that only parents and not students should receive information about grades. Six believed that students were the primary audience, and the remaining seven believed that parents and students should be equally informed. Four teachers specified that "next year's" teachers or colleagues should be included as primary audiences, in addition to students or parents. Other audiences—for example, administrators or future employers—were not identified as primary by any of the teachers.

To illustrate that parents should be the primary audience, two teachers felt that students already "knew where they stood." To illustrate that students should be the primary focus of outgoing information, one teacher commented that "students should take possession [of their learning]." A second teacher added, "They [students] are the ones who benefit from knowing how they [are] doing." Of the six teachers who believed that students should be the primary audience, four stated that although information should be directed to them, "we [teachers] aim for the parent," instead.

In addition to identifying primary audiences, the teachers were asked what they believed these audiences should know or would want to know as conveyed by the grades used. Eight teachers gave general responses for what students might want or need to know. Typical responses were "how well they [are] learning," "where they're at," and "whether [student] performance is satisfactory." Two teachers thought that students should know whether they mastered or knew the required material, and four teachers thought that students should know what tasks were graded and how grades were arrived at. Two teachers said that students would want to know if they were improving, one suggested that students would already know where they stood, and two teachers did not provide comments.

The teachers were more specific as to what they thought parents should know or want to know. Whereas seven teachers thought that parents needed or wanted information in terms of grade-level expectations or their child's standing relative to the curriculum, six teachers felt that parents needed or wanted information concerning how students compared to others. Interestingly, not one teacher indicated that norm-referenced information was necessary for students themselves to know. Two teachers thought that parents should know how grades were determined, and three believed that it was important to discuss a student's ability and what might be realistic expectations for their child when considering future programming.

Although the numbers are small, it is interesting to note that more teachers felt that students should be the primary focus of communication about grades, but the information that students would receive from their teachers appears less specific relative to what parents would receive.

Form of Communication

When specifically asked to rank the usefulness of five symbol systems commonly found on report cards (Question 5d), the majority of the teachers experienced difficulty in ranking them. They displayed a prolonged hesitancy when attempting to answer or offered to retry the exercise. They felt that none of the symbols satisfied all of the grading purposes and suggested that choice was often influenced by their perceptions of what would be most meaningful to parents. For example, five teachers chose symbols which communicated the 'relative' level at which a student was performing—at, above, or below grade level—and suggested that this system would be meaningful to parents and that the parent would "know where the child was at." One teacher stated that this option was "the most clear and the least misunderstood by parents." In contrast, four teachers stated that, even though parents wanted to see a certain symbol, they believed that parents' choices should be disregarded in favor of systems that the teachers felt were more useful.

Although the teachers were not able to rank the five symbol systems provided, they were able to identify the one or two symbols that they believed to be most useful. As shown in Table 3, letter grades were the clear choice of 12 teachers, and of those, 7 stated the need for descriptors in order to achieve meaning. Six teachers favored grading by choosing the options *at, above, and below* grade level, and four teachers chose the words *excellent, good, fair, poor.* Three teachers picked the *percentage* grading system. Only one favored the *pass/fail, satisfactory/unsatisfactory* option.

Table 3

Teachers' Choices of Grading Symbols

Options	n[a]
A, B, C, D, F	12
At, above, below	6
Excellent, good, fair, poor	4
Percentage (%)	3
Pass/fail, satisfactory/unsatisfactory	1

[a]Total possible responses = 19.

To describe the reasons for the teachers' choices, the five options are split into two groups. Group 1 consists of letter and percentage grades; Group 2 includes the word descriptors. Though some teachers found two or more equally useful symbols to use, in all but one case the choices came from the same group.

Of the 12 teachers in Group 1, 9 chose the letter symbols and 3 chose percentage. They chose these symbols because they are easy to use, they denote a specificity of meaning, and the negative connotation of words such as *poor* or *fail* is avoided. One teacher, citing efficiency, commented that letter grades are "easy to use [because they] have corresponding percents built in, . . . [and one] can easily see how you got A, B, C." Another teacher indicated that "words are abstract; . . . [letter grades] are broad enough." A third said, "Words can be negative." A fourth teacher felt that "[we] use words when we explain grades."

The seven teachers in Group 2 who preferred word symbols stated that words communicate more fully and seem to be more appropriate for elementary-school-aged children and their program of studies. One teacher stated that "a percent is meaningless. . . . What does it tell?" Another commented that "we [students] don't learn for ABCs. . . . I may have gotten a C, but I may have learned a great amount." A third said that "at/above is tied to curriculum expectations."

In addition to the variety of reasons that teachers reported for using symbols, they felt that the usefulness of a grading system often is dependent upon what is to be reflected in and communicated by a grade. For example, 14 of the teachers viewed the grading symbols in relation to the achievement domain. The remaining five suggested using symbols for effort, as well as for achievement. Two of these teachers were in favor of using common symbols for both achievement and effort; the remaining three indicated that they would use different symbols; for example, E, S, or N for effort and A, B, C, or D for achievement.

The choice of symbols did not appear to be related to the grade level at which teachers taught. Contrary to expectation, there were no visible differences between Grade 4 and Grade 6 teachers when choosing a word, letter, or number grade system. Further, Grade 6 teachers were not influenced by the grading system used at the next school level, junior high, where percentages are used.

Student Characteristics Included in Grades

Though there was variation as to which specific student characteristics should be considered when forming grades, all teachers believed that both achievement and all or some nonachievement variables identified (Question 4a) should be included (see Figure 6). As shown in Table 4, all of the teachers believed that achievement, effort, and work habits should be considered and included in the formation of a grade. Nine teachers believed that all seven of the nonachievement factors-effort, aptitude, improvement, work habits, attitude, class participation, and co-operativeness--should be considered in addition to achievement. Of the remaining 10 teachers, 7 believed that aptitude or potential is not appropriate, 3 felt that improvement has little or no value in grading, 3 felt that co-operativeness should not be included, 1 was unsure whether co-operativeness should be included, and 3 disregarded or were unsure about one or both of the remaining attributes of attitude and class participation.

Table 4

Teachers' Beliefs About What Characteristics Grades Should Include

	Student characteristics							
Resp	Ach	Eff	Apt	Imp	WH	Att	CP	Coop
1	*	✓	x	✓	✓	✓	✓	✓
2	✓	*	✓	✓	✓	✓	✓	✓
3	*	*	x	x	✓	✓	✓	✓
4	*	*	✓	✓	✓	✓	✓	✓
5	✓	✓	✓	✓	✓	*	✓	✓
6	*	✓	✓	✓	✓	✓	✓	✓
7	*	✓	✓	✓	✓	?	?	?
8	*	*	✓	✓	✓	✓	✓	✓
9	*	*	x	✓	✓	*	x	x
10	✓	✓	✓	✓	✓	*	✓	✓
11	*	✓	x	✓	✓	✓	✓	✓
12	✓	*	x	[illegible]	✓	✓	*	✓
13	*	*	x	✓	✓	✓	✓	x
14	✓	✓	✓	x	✓	*	✓	✓
15	*	✓	✓	x	✓	x	x	x
16	*	*	✓	✓	✓	✓	✓	✓
17	*	*	✓	✓	*	✓	✓	✓
18	*	✓	✓	✓	✓	✓	✓	✓
19	*	*	x	✓	✓	✓	✓	✓
Totals	19	19	12	16	19	17	16	15

* = Included and most/equally important
✓ = Included
x = Not included
? = Unsure

Resp = Respondent
Ach = Achievement
Eff = Effort
Apt = Aptitude
Imp = Improvement
WH = Work habits
Att = [illegible]ude
CP = [illegible]participation
Coop = [illegible]erativeness

In addition to the inclusion of different student characteristics in a grade, there were differences with respect to the emphasis the characteristics received. Six teachers felt that achievement alone is the most important student characteristic to include in a grade, and six teachers rated achievement and effort equally important. Three teachers felt that attitude is most important. One teacher rated achievement, effort, and attitude equally important; and one rated achievement, effort, and work habits equally important. One teacher felt that effort and class participation are most important, and one felt that effort alone is most important.

When asked how to combine the information about the various student characteristics that they considered to form grades (Question 4b), the results were mixed, though all but one teacher favored the use of two grades, generally one grade for the achievement component and a second grade for the nonachievement component. As indicated in Table 5, 12 of the 19 teachers indicated that achievement should be graded separately; an additional six suggested that achievement should be combined with one or two other characteristics: improvement, potential, and/or class participation. All four of these variables have an underlying 'cognitive' basis, in contrast to the remaining student characteristics considered, suggesting that these teachers were differentiating between cognitive and noncognitive characteristics.

The nonachievement student characteristics were generally combined in one of two ways. Eight teachers preferred to grade all of the nonachievement variables together, referring to it as an effort grade. Eight others believed that effort should be graded alone or in combination with one or two other characteristics: work habits, attitude, class participation, and/or co-operativeness. One teacher favored a single holistic grade for all student characteristics, one felt that each student characteristic should be graded separately, and one was unsure as to how to combine the various characteristics into grades. Of the second subset of eight teachers, five indicated that

Table 5

Teachers' Beliefs About Assigning Grades to Student Characteristics

Resp	Ach graded	Ach plus graded	Effort graded	Effort plus graded	Non-ach graded	Non-ach sep. report	Non-ach sep. graded	Global mark
1	✓				✓			
2	✓		✓			✓		
3	✓				✓			
4		✓I		✓WH		✓		
5	✓		✓			✓		
6								✓
7	✓		?			?		
8	✓				✓			
9	✓				✓			
10	✓						✓	
11		✓I/CP			✓			
12	✓		✓					
13			✓			✓		
14		✓P/CP		✓WH/A				
15		✓P		✓WH		✓		
16		✓I/WH		✓CP/CO				
17	✓				✓			
18		✓I			✓			
19	✓				✓			
Totals	12	6	4	4	8	5	1	1

✓ = Choice of grade assignment
? = Unsure

Resp = Respondent
I = Improvement
WH = Work habits
CP = Class participation
P = Potential
A = Attitude
CO = Cooperativeness

they would favor reporting the remainder of those student characteristics not considered for a grade by some other means; for example, in written comments.

Teachers focus on different student characteristics and combine them in diverse ways according to what they believe to be validly represented in a grade. It appears that teachers favor a two-grade system which may be attributable to the district's reporting system which has a built-in mechanism for separating grades. Beyond that, however, several teachers found the task of choosing and combining student characteristics challenging and labored at length in identifying the variables and how they should be combined to obtain a grade or grades which appropriately represented a student.

Context (Internal Environment)

As shown in Figure 6, the teachers described three conditions internal to the student which they need to consider when formulating a student's grades. These conditions are (a) the innate ability of students, (b) motivation and other characteristics of students, and (c) negative consequences that are potentially harmful for students. Together, these three cluster themes form some of the mediating variables discussed previously and illustrated in Figure 4.

Assigning Grades According to Ability

When the issue of differences in student ability was addressed, the teachers' responses were varied. Asked whether different 'ability' groups should be graded according to different criteria (Question 6a), 10 teachers responded "yes," 7 said "no," and 2 were "uncertain."

Further, of the 10 teachers who responded "yes," 5 indicated that provisions should be made for special-needs students. Two of these teachers were special education teachers and made their comments with regard to only special-needs

students. The remaining three teachers appeared unwilling to extend special provision to other ability groups. For example, one of the three teachers who would make special provisions only for special-needs students commented that "they [are] identified as special needs so should be assessed this way." Another commented that "it was unfair to parents and children [of regular classes] to have separate criteria."

In contrast, the remaining five teachers who responded "yes" suggested that all differences in ability should be considered when developing a grading system. One of these teachers commented that "if students are told that 'you are not succeeding,' that creates problems too."

To illustrate the reasoning of the seven teachers who would not make provisions for ability when developing grades, one teacher commented that teachers should use one set of criteria because "they [students] must be judged according to a standard." Another teacher elaborated that others "need to know how they [students] stand against the population at large." A third suggested that children become unfairly labeled "I'm a crow or I'm a robin" and that it is preferable to avoid "pigeon-holing kids."

Even though these seven teachers provided comments like those presented, they nevertheless made qualifying statements. One recognized that "judging quick versus slow kids is not fair." Four others commented that although consistency of grading practices (and therefore meaning) is of primary importance, they specifically use other means to qualify an otherwise literal meaning of a grade. Two teachers compensate for low achievement by awarding a higher effort grade. One teacher said that even though a standard is used, each child has a different combination of strengths or weaknesses and that these are focused on instead, thus giving evaluation an individualized nature. One teacher said that even though the achievement criteria remain the same, it is "the way of getting there" which differs; often adjusting time requirements for slower working students helps them to achieve. Though the

discussion of modification of grading practices is elaborated upon in the next section, this discourse provides evidence that, contrary to what they reported, teachers modify their grading practices under extenuating circumstances.

In a later question (Question 8), when the teachers were asked to identify specific difficulties that they experience during grading sessions, eight teachers isolated the problem of grading low-ability students. Of these eight, three had previously stated that ability should not be given consideration when developing grades, and five (two of whom considered ability in special-needs students only) stated that it should be considered. This finding suggests that the consideration of ability when forming grades is a complex and often contradictory issue and that perhaps teachers apply their beliefs in terms of making decisions more on an individualized basis.

Motivation and Other Characteristics of Children

As previously indicated, although 16 teachers believed that grades would influence their students, they later qualified this hope, commenting that, in reality, how influential grades are depends upon four factors: intrinsic motivation, potential for achievement, a competitive nature, and a goal-setting orientation (see Figure 6). As indicated in Table 6, 13 teachers felt that grades motivate some students but not others (Question 3b). Five teachers felt that grades motivate those students who are intrinsically motivated at the outset or who have "inner drive." Four teachers believed that grades are influential for those students who possess a high potential for achievement, say, "the top 20%." Five teachers submitted that a competitive nature is motivating; for example, for the student who "boasts about As"; whereas four teachers felt that students who are goal-oriented by nature are motivated by grades; for example, when a student asks, "How much does it [the mark] count?"

Table 6

Teachers' Perceptions About the Influences of Grades

	n[a]	Yes	No	Qualified
Grades are meant to influence students	19	16	2	1
Grades motivate students	19	4	2	13

[a]Total possible responses = 19.

The teachers suggested that those students who do not characteristically belong to the groups of students described in the previous paragraph are not motivated by grades. Five teachers noted, for example, that low-ability students are not motivated by grades. One teacher thought that grades are "frustrating," and another said that grades are "irrelevant to bottom students." Another teacher stated that they are not beneficial for the "C and D students" who are in a "slump" because their grades do not indicate that progress is being made. Five teachers believed that students of this age bracket are not goal oriented, that at this age the students are not "cognizant of long-term goals" and "do not see the long-term reward."

The teachers appeared to believe, at least in part, that because of the very nature of children and their vast differences, they could not be expected to adopt a "one-size-fits-all" policy of grading and that distinct, and often opposing, characteristics within children have to be taken into consideration. This supports the earlier finding that, for some teachers, ability is a factor to be contemplated when grades are assigned.

Potential Harmful Consequences for Children

All but two of the teachers believed that grades may negatively influence a student, particularly the low achiever. This individual theme constitutes the third and final cluster theme in this section.

The teachers cited several different forms of negative influence (Question 3d): loss of self-esteem, pressure or other repercussions from parents, lack of advancement, and behavior problems. One teacher commented that a student could get "discourage[d] . . . [with] loss of self-esteem." A second asked, "What will marks do for [the] child's ego?" A third teacher elaborated that a child "who finds school too difficult and [is] afraid to go home with a negative report card . . . or a child whose parents want extremely high performance at all times" is at risk. A fourth teacher proposed that "often a D student works hard but is denied the opportunity" for advancement. Lastly, one teacher stated that "inappropriate behavior [occurs] as a result of frustration."

When invited to provide additional comments about the influences that grades might have on their students (Question 3e), 17 teachers did so. Seven teachers discussed the unfairness of comparison or competition. Though the majority of these teachers were concerned about the influence of grades on low achievers, three teachers added that a high degree of competition and comparison often exists for students of high or average ability and that this affects them in negative and "unhealthy" ways. One teacher stated that such students do not have a "rounded and balanced life" because they are often expected to overachieve. Five teachers expressed concern about the high expectations held by parents and the rewards these parents provided if their child(ren) achieved these expectations. Eight teachers had concerns about the grade system itself, because it seems to foster comparisons or competition and adherence to extrinsic motivators. For example, one teacher commented, "If kids come to school to get good grades, that's a problem." Another

objected, "Elementary school is too early to be concerned with ABCs" and added that competition should be "saved for later on."

Modified Practices of Teachers

Taking into account the role that context plays when teachers are involved in their grading practices, the results indicate that teachers modify or alter their grading practices as a way of 'staying true' to their beliefs. How grades are assigned or should be assigned often deviates from their beliefs about grades. As a final section in this chapter, the results indicating how teachers modify their grading practices is described.

Sixteen teachers outlined an action that they would undertake to avoid an unpleasant consequence and/or alleviate a concern. From these responses, two individual themes which constitute the final cluster theme were identified: Teachers 'educate' parents and students about their grading systems, and they 'change the rules' when assigning grades.

Educating and Clarifying

The teachers suggested that it was necessary to better inform parents about grades. Three teachers indicated that it is necessary to explain their grading system to parents. For example, one teacher stated that "[I] need to explain growth because a mark or a percent may not always show that." Three teachers mentioned that parents need to be made aware of realistic expectations for their child "because parents need to know what is a good grade for their child." For example, one teacher suggested that teachers should inform parents of "strengths and weaknesses of individuals."

Teacher Adjustments

Four teachers emphasized the advantage of the two-grade system and the role that the effort grade has within this system. One teacher suggested, "[Raising] the effort grade would create success [for the student]." Another commented, "[Students] will get a grade so that they feel that the effort has been worthwhile." Four teachers discussed how important it was to create opportunities for students to experience success. For example, one teacher said that "all students should be given the opportunity to show excellent work." A second added that having the student do only a portion of the required work would "build in a chance for success."

Three teachers established that, when grading, they take a more individualized approach. They commented that doing so allowed them to better communicate the strengths or personal growth of each student.

Two teachers mentioned "supporting" the student by "working extra hard on those students to help regain confidence in their abilities." This support included reteaching concepts, communicating to students why they did poorly and offering suggestions for improvement, and teaching students how to set goals.

Though teachers indicated concerns about the negative consequences that grades might have for children and stated that they would take alternative action, when asked specifically if there were situations where they would 'bend the rule' and consider other factors when assigning a final grade (Question 4d), responses were mixed or indecisive. Nine teachers reported that there are circumstances that would warrant "fudging" a grade. Eight teachers reported that they would not change the grade and two others were unsure. This finding illustrates the complexity and complications of decision making in the presence of contextual factors. What is apparent is a contradiction between the desirability of consistent practice and the need to accommodate individual student situations.

Summary

The functions of grades as teachers understand them have been presented in this chapter. These findings were supported by the teachers' beliefs that grades are intended to serve useful purposes. Teachers recognize that grades should be used to communicate effectively to a variety of audiences. They are able to make firm choices in terms of symbol systems for communicating the grades. Though teachers attach different importance to the various student characteristics and combine them in various ways, a two-grade system reflecting a cognitive and a noncognitive component is most common.

However, teachers recognize that, in the presence of differing student abilities, differing student motivational or other characteristic traits, and potential harmful consequences of grades for students, it is not possible to have a simple, straightforward system of grades that validly communicates student behavior and performance.

Chapter VI

Teachers' Beliefs About The Process of Grading

Overview

The findings in this study have been reported in a particular sequence. This sequence reflects the overall finding that teachers appear to have general meanings for a grade or the process of grading (Chapter IV). Further, they understand the functions of grades, but recognize that there are contextual factors which, when accommodated in their grading practices, result in additional or different meanings of grades as compared to those defined in the absence of these mediating factors (Chapter V). In this chapter, the results and discussion related to the processes of grading are interpreted within additional contextual situations, including issues related to factors external to the students' environment.

The Process of Grading

As shown in Figure 7, the teachers interpreted and described the process of assigning grades in four ways: (a) implementation of a frame of reference for assigning grades, (b) procedures for determining grades, (c) identification of measures to use for establishing achievement grades, and (d) procedures for combining scores from these measures.

Further, when formulating grades, teachers accommodate external contextual factors as they occur in the real workplace. Issues regarding the applicability of consistent grading practices from the perspective of various levels of decision making and the problems of weakness in the system are taken into consideration. Their influences on grading processes, as reflected in the teachers' responses, are discussed last. Further discussion of each of these cluster themes in terms of their component individual themes follows.

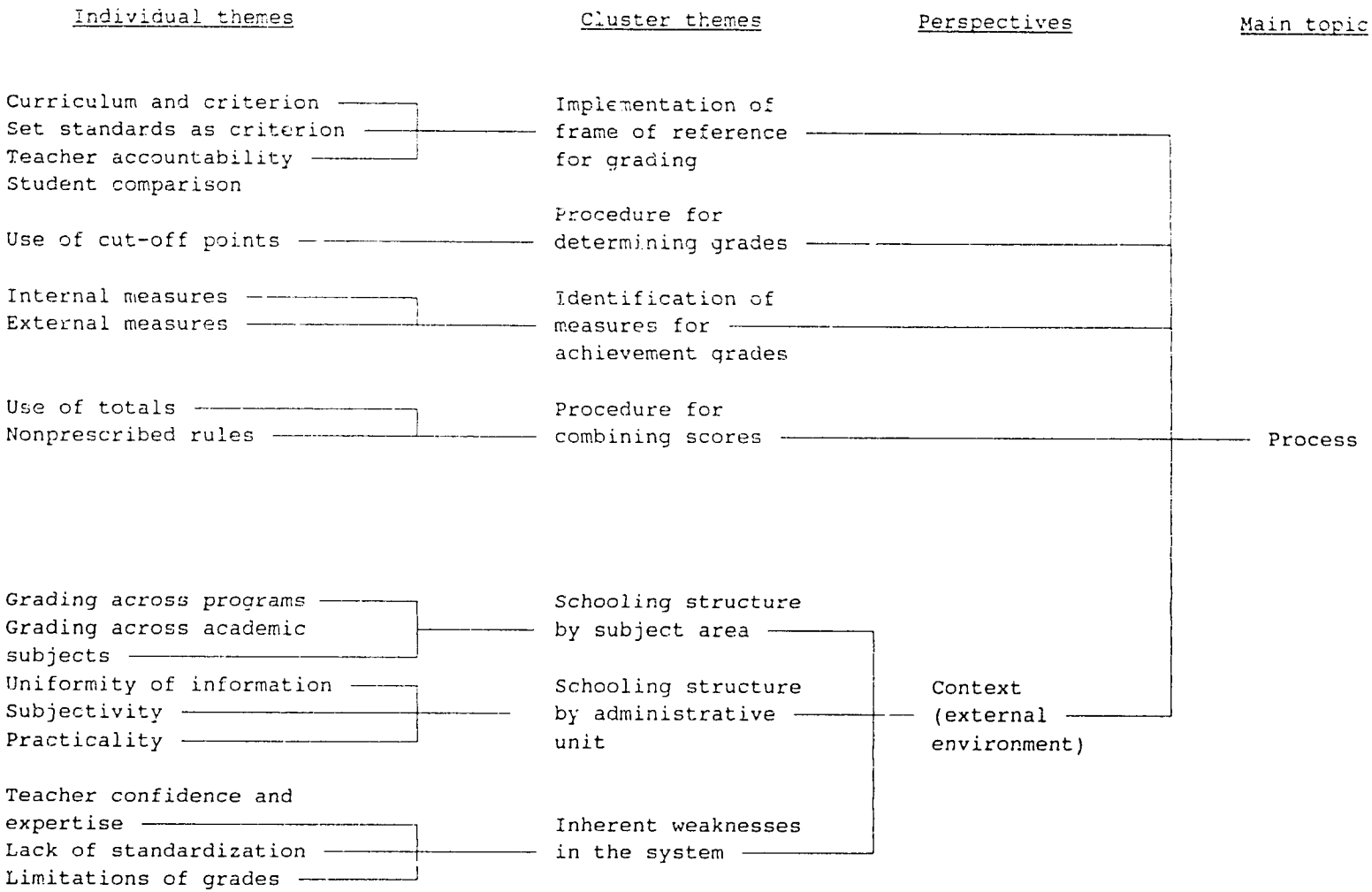

Figure 7. Progression of analysis identifying the process of grading.

Implementation of a Frame of Reference for Grading

It will be recalled from Chapter IV that 17 of the 19 teachers identified that grading involves making a comparison relative to a particular interpretative framework. When teachers were asked to choose the most appropriate frame of reference to use when assigning grades to their students (Question 5a), 13 believed that a criterion-referenced framework (CRF) is the most appropriate, 2 believed that a norm-referenced framework (NRF) is the most appropriate, and 1 believed that personal growth or an ipsative-referenced framework (IRF) is the most appropriate. The three remaining teachers believed that a combination of methods is most appropriate when assigning grades (see Belief column, Table 7). When asked to provide reasons of choice, an analysis of the teachers' comments reveals that the first cluster theme is comprised of four individual themes (see Figure 7).

Table 7

Teachers' Beliefs and Practices About Grade Assignment

		Practice (n^b)			
Belief	n^a	CRF	NRF	IRF	Comb.
Criterion-referenced framework	13	11	1	0	1
Norm-referenced framework	2	0	2	0	0
Ipsative-referenced framework	1	0	1	0	0
Combination	3	2	1	0	0

[a]Total possible responses = 19.
[b]Total possible responses = 19.

Of the 13 teachers who believed that the CRF is the most appropriate, 4 indicated that the standard to use should be defined in terms of the curriculum set out by Alberta Education. They supported the adoption of the set curriculum because the curriculum is what is to be taught. One teacher put it succinctly: "If Alberta Ed sets

out criteria, this is what students need to know to have mastery at this particular grade level, and if my students have met that . . . [teacher gestures to emphasize point]." Five other teachers submitted responses indicating that some form of set standards or levels should be used by the majority of teachers. One teacher explained that it is "important for students to be aware of the fact that we do have set standards, . . . [and students] should be recognized by them," and a second added that "[parents are] looking for a standardized idea of what is happening throughout the system."

Five teachers, one of whom was included in the above discussion, were cognizant of their own responsibilities. Because they are accountable to others, these teachers believed that the CRF is the best available option to fulfill their roles as evaluators of students. One teacher paraphrased: "My job is to have students master the Grade 6 curriculum (standards and behavioral objectives), so I look at those objectives and try to have students meet them."

Of the two teachers who believed that a child's performance relative to peers (NRF) is the most important, one commented, "I don't know what a perfect Grade 4 student is, but I can look at my group and see who is doing better than most, . . . when out in the real world that is how you are judged." One teacher felt that any comparison to external criteria is "too frustrating." He felt that it is "fairer" and more "realistic" to have a standard for each child (IRF). In practice, each child's scores are taken on their "own merit." A final grade is also dependent on the amount of effort given or the amount of ability that the student possesses. This teacher's practice is consistent with his view of the holistic child where the child should be assessed in global terms with one grade.

Of the three teachers who choose a combination of frameworks, two suggested that the most appropriate framework is subject dependent. As one teacher observed, "There's always a standard [for math], but LA is very subjective." This indicates that for math a CRF is the most suitable method, but language arts is more individualized,

and growth (IRF) is viewed as more suitable. The third teacher believed that all methods should be viewed simultaneously prior to making a decision: "I compare with my expectations, other students' marks, but also take into consideration this mark for this person."

Procedure for Determining Grades

In order to determine the consistency between the teachers' beliefs and their practices concerning frames of reference for grading, each teacher was then asked to demonstrate the assignment of grades from a hypothetical set of class scores (Question 5c). The procedure was then categorized by the researcher as one of NRF, CRF, IRF; or a combination. The results of the analysis are presented in the practice column of Table 7. For convenience of comparison, the left column of Table 7 contains a summary of the teachers' beliefs.

Of the 13 teachers who previously identified the criterion-referenced framework as the most appropriate method to use when forming grades, 11 did so and chose a preset standard for determining cut-off scores. Of the two remaining teachers, one uses a combination of a CRF and an NRF approach, stating that both standards and empirically derived methods are equally important. As suggested by her comment, "I have to value everybody's work so [I] have to have some spread," she considers curriculum expectations but feels that it is important to find finer distinctions among students' abilities. The second, although purporting to use standards for judging students' work, appeared to be more strongly motivated to grade on an empirically derived basis. She commented that she does not "want them [students] all to get A's." Furthermore, there was evidence that she used natural breaks post hoc when determining cut-off points.

Both teachers who believed that an NRT approach was most appropriate use this approach when grading and chose the empirically derived method for determining

cut-off scores. One teacher initially said that an ipsative framework is most important; however, as shown in Table 7, he adopts an NRF approach and chooses empirically derived cut-off scores, but "only as a guide."

Of the three teachers who believed that a combination of two or more frames of reference is most useful for assigning grades, two use the CRF approach and chose the preset standard when determining cut-off points. The remaining teacher admitted that although he uses a combination of frameworks, in the final analysis he likes to "see the students in context" and uses an NRF approach.

Although five teachers use the NRF approach to determining cut-off scores for determining grades, no one chose a preset distribution for grading purposes. A number of teachers questioned the validity of tests or felt that norming is not appropriate for them because "we deal with a nonstandardized situation in the class[room]."

Teachers appeared to be consistent in their beliefs about using a CRF when assigning grades. This is supported by the earlier finding in Chapter IV that the majority of teachers refer to judging students on a criterion-referenced basis. This interpretation is supported further by the emphasis that teachers gave to using preset standards in their practice of determining cut-off scores.

Identification of Measures for Achievement Grades

In order to understand better the components that teachers include in a grade, two specific questions related to this issue were asked. The first, discussed here, asked teachers to choose from a list of prescribed measures those which they felt were useful for determining achievement grades (Question 9). The second question, discussed in the following section, asked teachers how they combined the scores from these measures prior to assigning a final grade (Question 5b).

As displayed in Table 8, patterns showing the use of measures clearly exist. Some measures are chosen more frequently than others, and some measures are given more emphasis than others. For discussion purposes they are divided into two groups: those measures created internally by teachers and those published by outside educational agencies (e.g., standardized tests).

Internal measures. Of the 18 teachers for whom data were available, all use class assignments and projects to determine a student's grade (see Table 8). This was followed by a strong emphasis on paper-and-pencil tests (n=17) and observations (n=15 . There was a lesser emphasis on oral questioning (n=11). Although 10 teachers felt that intuition plays a part in achievement grading, this issue was contested. For example, one teacher commented that "intuition plays no part in achievement"; and a second emphasized, "Definitely not." In contrast, three of the teachers who indicated that intuition plays a part in grading pointed out that subjectivity is a necessary part of teachers' grading processes. Half of the teachers (two of whom believed that it was most important) included homework as a component of grades.

External measures. Although all 18 teachers for whom data were available indicated that they use the results from standardized tests (e.g., Gates-McGinitie, Canadian Test of Basic Skills [CTBS], Canadian Cognitive Abilities Test [CCAT]), only five reported that they combine these results with the results from their own measurements to formulate a grade (see Table 8). An additional teacher reported the standardized results separately.

Twelve teachers (including two of the teachers who incorporated standardized test results in their grades and the one teacher who reported them separately) remarked that the standardized tests serve a comparative role: Results from

Table 8

Teachers' Identification of Measures for Achievement Grades

	Internal measures						External measures (standardized tests)			
Resp	A/P	PPT	O	OQ	I	H	IG	RS	NIG	CT
1	*	✓	✓			✓		□		□
2	*	*	✓	✓	✓	✓	□			
3	*	*	✓	✓		✓	□			
4	*	✓	✓	✓					□	
5	*								□	□
6	*	✓	*	✓	✓	✓			□	
7	*	*	✓	*	✓		□			□
8	*	*	✓	✓	✓	*	□			□
9	*	*				*			□	□
10	*	*	*	*	*				□	
11	✓	✓		✓					□	□
12	✓	✓	✓		✓				□	□
13	*	*	✓			✓	□			
14[a]										
15	✓	✓	✓	✓	✓				□	□
16	*	*	*	✓	✓	✓			□	□
17	*	*	*						□	□
18	✓	✓	✓	✓	✓				□	□
19	*	*	✓		✓	✓			□	□

* = Included and most important
✓ = Included
(blank) = Not included
[a] = Missing data
□ = Uses of standardized tests

A/P = Assignments/projects
PPT = Paper-and-pencil tests
O = Observations
OQ = Oral questions
I = Intuition
H = Homework

IG = Included in grade
RS = Reported separately
NIG = Not included in grade
CT = Comparative tool

teacher-made measures are compared with the results from the standardized tests. The standardized tests provide a "backup" and "reassurance," are "my yardstick," and furnish "justification for my own tests." Five teachers noted that they use the standardized test results "for diagnostic purposes." Two teachers suggested that these results indicate a general ability or potential in students, but added that they should not be included in a grade.

In addition to the standardized tests discussed above, teachers were asked to provide comments on the use of districtwide tests. Half of the teachers felt that they were not useful, citing reasons of invalidity: Tests "do not match the curriculum," "do not change [from year to year]," "[are prone to] teaching to the test," and "[are] too traumatic." Of the five teachers who said that they would like to use district tests, three noted that the results are not returned to the schools in time to be useful for reporting purposes. One teacher stated that district tests that are used primarily as surveys are a waste of resources but that they could be useful. One teacher felt that these tests "are matching the curriculum better now" and have the potential to become more useful with time.

It is clear that teachers prefer internal measures for determining grades and reporting student performance. Interestingly, in later findings teachers indicated that standardized measures would be an asset to their grading practices.

Procedure for Combining Scores

When teachers combined the scores from the measures they use to formulate a grade, the procedures are not consistent across teachers, and in some cases are not clear. Nine teachers reported working with numerical scores which they total prior to assigning grades. Of these nine, four specified that they "quantify" as much as they can. They felt that this is a more efficient way to account for the numerous performances that occur over the course of the reporting period.

An additional five teachers commented that they use other methods of combining scores in addition to adding numerical scores. Of these five, four commented that the use of numbered totals depends on the subject. For example, one teacher said, "[Language arts] gets holistic impression. . . . [I use] anecdotes, gut feeling, samples [of work]." A second teacher added, "For social studies [I] can't assign a number [but] will record a general impression."

Five teachers did not use numerical scores at all. It was evident that these teachers often combine a number of tasks in idiosyncratic ways from which evolves a grade; for example:

> I subjectively give A-B-Cs, then say, "Does it fit with that person on a daily basis?"
>
> [I use] internal averaging of daily work [mental recordkeeping].
>
> [I] look at the whole unit experience [global mark].

Context (External Environment)

As shown in Figure 7, the teachers considered outside influences or contextual factors that impact on teachers' grading practices. Three conditions were identified from their responses: (a) schooling structure by subject area, (b) schooling structure by administrative unit, and (c) inherent weaknesses in the system. These cluster themes form the external mediating variables introduced earlier in Chapter IV (see Figure 4).

Schooling Structure by Subject Area

When teachers were asked if subjects with different program objectives should be graded differently (Question 6b), their responses revealed that they can be divided into two groups. One group, consisting of eight teachers, appeared to hold the view that academic and nonacademic programs and subjects should be graded using a

common grading system. The second group, consisting of 11 teachers, suggested that this common view is not appropriate, and that separate grading systems should be employed (see Table 9).

Table 9

Teachers' Beliefs About Consistency of Grading Practices

Condition	n[a]	Yes	No	Unsure
Across programs	19	8	11	0
Within academic subjects	19	15	2	2
Within school/division level	19	19	0	0
Within district	19	14	3	2

[a]Total possible responses = 19.

Grading across programs. Of the eight teachers who indicated that programs with different objectives should be graded consistently, four believed that the nature of prescribed curricula allows for grading practices to follow a consistent procedure. One teacher remarked that "every subject has objectives, and we have to decide if kids have met them. . . . The achievement [of performance programs] is a different sort, . . . [but] you look at the final product for completion of criteria." A second teacher remarked that "there are measurable benchmarks in all subjects."

Five teachers (two of whom belonged to the same subset above) suggested that competent teachers who know and understand the curriculum should be able to grade consistently across all subject areas. For example, one teacher said that "if you're teaching to a set of curricula, the onus is on the teacher to get skills for doing the job [of grading]." A second teacher stated that "I can grade music the same as other subjects," and a third commented that "if teachers say they can't grade [art], then perhaps they [are] not teaching it right."

In contrast, of the 11 teachers who indicated that it is not possible to use a common grading system, 6 felt that a student's participation was the primary focus of performance-based programs such as music or physical education, whereas a level of achievement is more relevant in programs such as reading, mathematics, social studies, and science. One teacher stated that "music should be only effort and participation." A second teacher proposed that "students see physical education as a lifelong attitude thing, . . . [and] we want to establish a positive attitude for kids." A third teacher said, "[We want kids] to risk a little and not focus on academics." The remaining five teachers did not elaborate with reasons.

Grading across the basic academic subjects. There was greater agreement among the teachers regarding the appropriateness of consistent grading across the four core subject areas (Question 6c) than there was regarding the appropriateness of consistent grading across the program areas. As shown in Table 9, 15 teachers agreed that grading practices should be consistent across the core subjects, 2 disagreed, and 2 were unsure. The four teachers who disagreed or were unsure about similar grading practices across the core subjects expressed difficulty with grading math and language arts in a consistent way.

Schooling Structure by Administrative Unit

As shown in Table 9, all 19 teachers agreed that grading practices should be consistent within a school or within division levels in a school (Question 6d). At the district level, 14 teachers indicated that the same process should be followed (Question 6e), whereas 3 teachers disagreed, and 2 were not sure. The teachers who endorsed school-based consistency but not districtwide consistency felt that the "uniqueness" of schools should have priority in terms of meeting the needs of individual neighborhoods, because "each community has its own needs." This

philosophy is congruent with the stipulation in the definition of grade that its meaningfulness is dependent upon its shared understanding within a community.

The most frequently cited reason for uniform information was to increase the understanding of grades (n=15), particularly by parents. One teacher commented, "If what we are communicating to parents is to have meaning, we have to be consistent." Another suggested, "[There are] too many differences between teachers, [and] parents have problems understanding grades." A third teacher, cognizant of public awareness, added, "We need to show them [the public] we're consistent and we're all working in the same direction." A fourth recognized the advantage of uniformity when students move from one school to another and added "for ease when children move."

Although the teachers believed that it was important to share uniform information, there was some feeling that this would be difficult to achieve because of teacher subjectivity and the issue of practicality. Six teachers noted that although "a greater universal" is ideal because "the closer you get to continuity the sounder the basis," it would be difficult to achieve consistency because it would be impossible to rule out personal subjectivity and professional judgment. Two comments encapsulated this idea: "Each teacher has [his] own standards anyway" and "Each of us arrives at a certain point with different expectations."

Further, whereas two teachers felt that system- or districtwide consistency is imperative, five teachers thought that it would not be feasible or practical. One teacher, in favor of districtwide consistency, commented, "[We] should be sailing by the same chart." In contrast, a teacher who questioned the practicality of districtwide consistency of grading practices emphasized that a common policy is "too broad and [would] not say much of anything." A second teacher felt that feasibility is a problem and remarked, "It's hard enough to get consistency across schools." Two additional teachers questioned the wisdom of a systemwide policy, one suggesting that "it would

be difficult to be consistent; . . . there is so much personal stuff [in children] that would not be accounted for"; and the second added that a common policy "does not often tell the truth about what a child can do."

When the teachers were asked whether they were able to follow their own beliefs when grading (Question 8a), 16 of the 18 teachers for whom responses were obtained replied that they were able to do so. Two responded that they were mandated to grade according to a policy contrary to their own. This finding may explain, in part, why teachers strongly endorse a grading policy in an effort to achieve a high degree of consistency of meaning.

Inherent Weaknesses in the System

In many large organizations, the difficulty of implementing one's beliefs and philosophies is a common reality and a source of frustration. As indicated above, consistency is difficult to achieve because, in reality, a consistent practice "breaks down the larger the units." Fifteen teachers emphasized the weaknesses in the system that exist by definition in large organizations and that directly affect the meaningfulness of grades. As shown in Figure 7, the teachers' comments made in support of their thoughts were classified in terms of three individual themes: the level of teacher confidence and expertise, a lack of standardization within the district, and the limitations of grades.

Teacher confidence and expertise. Two teachers reported that in their experience of working with students' cumulative folders or report cards, they had not noticed a great discrepancy across teachers' judgments and felt that grades are consistent and truthful (Question 7). A third teacher said that her grades and those of her colleagues are accurate and that she believed that "teachers' grades are best for making decisions." In contrast, nine teachers expressed difficulty with grading

because they doubted their abilities. Their insecurity applied to both grading achievement and nonachievement, as seen in the comments they made:

> How do [I] grade L.A?
>
> Am I accurate in my judgments, especially effort?
>
> [Am I] judging growth or achievement level?
>
> [I find that] I do it [the process] many, many times before finalizing marks.
>
> [I have trouble] balancing the positive with the negative. . . . I don't know how subjective to be.

Lack of standardization. Nine teachers observed that because standardized tests or specific criteria for determining grades, and/or grading policies are not widely implemented across the district, consistency of meaning is compromised. The following are typical of the wide variety of comments:

> Unless they [the district] give us a criterion-referenced package to use, [I] can't see how . . . [consistency is] possible.
>
> [There is a] lack of standardized tests to give grades meaning.
>
> [There is] not enough breakdown of criteria to have consistency.
>
> [Audiences] don't know what grades mean across schools or programs.
>
> [It is] not known how grades are arrived at.
>
> [There are] too many ways of grading.

Limitations of grades. Eight teachers recognized the fallibility of the grading system. Five teachers doubted that grades are an accurate representation of a student's ability or that audiences interpret grading symbols in the same way. Two believed that the grade system is too "constricting" and that the "numbers are too precise." Two believed that by emphasizing the objectivity required for grades, teachers overlook other human considerations. They failed to see how a symbol could account for all of the "other [human] factors involved in the process of assigning grades."

To ascertain what teachers believed about the validity of grades, the teachers were asked whether they thought grades provided truthful, consistent descriptions of student achievement (Question 7). Ten teachers stated that they did not, four teachers thought that they did, and five teachers were unsure.

When asked whether evaluation could take place without a formal system of grading, 10 of the 17 teachers who responded believed that it could not. They suggested that a formal system is necessary for interpretation. One teacher put it aptly: "With a grade you can say, 'This is a "B" student.'" A second added, "You have to compare them [students] to something; . . . [evaluation is] meaningless." Although, as indicated earlier, grades were believed to be too brief, two teachers logically concluded that "often 'books' are written about children, and we still ask, 'Yes, but what can they produce?'" Furthermore, they submitted that "comments [are] too opinionated . . . [and that] at a glance grades give more info."

Of the seven teachers who felt that evaluation could take place without a formal system in place, three indicated that conferences would give parents an indication of how well their child was progressing. The remaining teachers discussed other nongrading methods, for example, "checklists," and "anecdotal comments."

Summary

The processes of grading as teachers practice them has been presented in this chapter. Teachers recognize and use an interpretative framework for assigning grades on a fairly consistent basis and are able to give reasons for choosing appropriate measures to use for grading. Though there was less unity across teachers when they combined scores, there was evidence that they had a plan, however subjective.

The teachers affirmed that contextual factors external to the students are at work during the grading process and that grading is dependent upon the subject area and level of school administration. They are acutely aware that limitations and

shortcomings exist and that they do not work in a perfect system. Furthermore, they do not overlook these conditions but accommodate them as a mediating influence in their grading practices.

Chapter VII

Summary, Conclusions, and Implications

Introduction

This chapter begins with a brief summary of the study, followed by a summary of the findings. After a discussion of the limitations of the study and an integration of the findings with the literature, the conclusion of the present study is presented. Following this are the implications for practice and research.

Summary of the Study

The main purpose of this study was to investigate teachers' beliefs about grades in terms of a number of specific issues related to their grading practices of upper elementary students. A second purpose of the study was to identify areas where discrepancies or possible deficiencies exist and where remediation might be appropriate.

A total of 19 teachers from a large school district in an urban center participated in the study. They had an average of approximately 20 years of teaching experience and were either currently teaching or had previously taught Grades 4-6.

A semistructured interview schedule composed of nine major questions formulated on the basis of the literature, measurement theory, and personal experience was used to gather information about the teachers' beliefs and practices. The questions covered a wide range of topics; for example, the influences of grades, student characteristics included in grades, interpretative frameworks assigned to grades, consistency of grading practices, and identification of measures, symbols, and procedures for assigning grades.

Each interview required approximately one-and-one-half hours of time. The interviews were tape-recorded and subsequently transcribed. The teachers' responses

and comments were processed using qualitative data analysis, following the procedures of Miles and Huberman (1984) and Constas (1992).

Summary of Findings

The teachers' responses were categorized into three interrelated topics: functions of grades, processes of grading, and a context in which functions were formulated and the processes were implemented.

Functions of Grades

The functions of grades, as described by the teachers, involve issues related to the purposes of grades and their communication.

1. Teachers hope that grades provide awareness, direction, and encouragement for their students and therefore believe that grades serve useful purposes. Ten of 17 teachers feel that a grading system is necessary for evaluating students.

2. They believe that letter grades are the most suitable forms of communication and that students and parents are the primary audiences for grades.

3. They favor a two-grade system, with one grade encompassing a cognitive component and the second a noncognitive component. The findings also reveal differences as to what specific student characteristics should be included in each of these grades. Of particular note are the differences in the way that effort is perceived and how it should be included in the noncognitive component of a grade.

Processes of Grading

The processes of grading as described by the teachers involve issues related to the implementation of a frame of reference for grading and the identification of measures that teachers use and the procedures that they follow in determining and assigning grades.

1. Thirteen teachers believe that the CRF is the most appropriate frame of reference for the assignment of grades, citing the importance of standards, of following the curriculum, and of showing accountability to the public.

2. Eleven of the above teachers who chose the CRF show consistency between their beliefs and practice by choosing a preset standard for determining cut-off points in order to determine grades.

3. Teachers use class assignments and projects, paper-and-pencil tests, and observations as primary measures for determining students' grades. About half of the teachers include oral questioning, intuition, and homework as part of students' grades.

4. Commercially published standardized tests are used by all teachers. A quarter of the teachers factor in the results of these tests with the students' final grade. However, the majority of teachers use standardized tests as comparative and diagnostic tools. Half of the teachers do not feel that districtwide tests are useful, although some stated that they would like to see district tests used more.

5. The results of how teachers combine scores prior to determining grades are mixed. Although half of the teachers reported using numbered totals prior to determining a grade, other methods appear to be mostly subjective in nature. It is apparent that the method of determining a final score is subject-dependent.

Contextual Factors Within Grading

Teachers recognize that they consider extenuating factors, both internal and external to the students' environment, that mediate between a given grading function or grading process and the meaning of a grade. To be compatible with their belief systems, teachers believe that the presence of these factors often necessarily warrants a change or a modification of certain practices.

1. When grading students of differing abilities, 10 teachers believe that these differences should be taken into consideration, either for all students or for special-

needs students. However, seven teachers believe that one set of standards should be used across all ability levels.

2. Thirteen teachers believe that grades motivate only those students who have high ability and a predisposition for self-motivation and competitiveness.

3. Seventeen teachers are concerned with various negative consequences that grades have on many children, and they seek to avoid these potential harmful consequences. However, contradictory results were given by seven teachers, who reported modifying a standard practice to accommodate individual students and who later reported that grades should not be changed because of extenuating circumstances.

4. Teachers perceive school programming as having both an academic and a nonacademic component and are fairly evenly split on the issue of whether a common grading practice should be used across all programs. The 11 teachers who oppose a common grading system cited that the broad objectives of these two types of programs differ, and the eight teachers who endorse a common grading policy believe that regardless of these differences, all objectives can be graded by a common system. However, there is a strong endorsement of grading consistently across the four academic subjects (n=15).

5. Although there is strong support for consistent grading practices within the district, teachers unanimously support consistent grading practices within individual schools or within each of the two division levels in a school, citing as a reason the need for consistency of interpretation by others. By the same token, teacher subjectivity and practical considerations are perceived to be incompatible with such a pervasive condition.

6. There is widespread belief that weaknesses of the system in the form of lack of teacher confidence and expertise, lack of standardization, and the limiting nature of grade systems mediates in and adversely affects a common interpretation of grades.

The relationship among the topics. The relationship between function and process is a fairly straightforward one. For example, when a teacher perceives the grading process as judging a student relative to one's peers, she attaches a particular meaning to that symbol which describes a particular student's performance. Each behavior that a teacher 'works through' results in a particular meaning for that teacher. There is a connection, an attachment between how a teacher does things and what that process means.

However, contextual factors, either internal or external to the student, create in teachers a deeper awareness of grades and their grading practices. Although certain questions during the interview were answered quickly and without reservation, the introduction of a particular context often changed a teacher's direction of thinking or revealed equivocal responses. Conflicts, complications, and contradictions arose as teachers realized that a particular context was often a mediating factor when making judgments and formulating grades. As indicated above, teachers often modify their grading practices in order to accommodate these contextual circumstances.

Limitations of the Study

The sample of teachers who voluntarily participated in this study may not be truly representative of the population of teachers. It is not known on what basis the participants volunteered, and why others, selected using probability sampling, did not. Therefore, the results should not be generalized beyond comparable groups of teachers in terms of geographical setting, type of district, age, experience, and levels taught.

Integration of Findings With the Literature

Introduction

Overall, there are marked similarities between the findings in the present study and those in past research; however, some notable differences are apparent. The following discussion is organized in terms of the three major topics identified in this study.

Functions

According to Hills (1981), the primary function of grading is to communicate information regarding academic achievement of students to a variety of audiences in some effective manner. Overall, the findings in this study are best represented by this view. The importance of grades in providing opportunities for self-reflection and self-evaluation (Manke & Loyd, 1991) is clearly evident in the responses of the teachers in this study. The teachers also appear to be similar to the student teachers in Barnes' (1985) study in that they include motivation and the 'sorting' of students as important purposes, in addition to the communication of achievement to parents and students. Contrary to the finding of Brookhart (1992), teachers do not generally view grades as a form of payment or something to be earned.

Parallel to the finding of Schulz (1993) is the belief of most teachers that some form of grading format is necessary for the evaluation of students. Though there is substantial favoring of word descriptors in both the above and present studies, the popularity of letter grades in this study is compatible with the findings of Austin and McCann (1992), Burton (1983), and Schulz (1993), but contrasts with the findings of Anderson and Bachor (1993). Teachers' beliefs that the primary audiences should be students and parents are supported by Burton (1983) and are reflected in the district policies reviewed by Austin and McCann (1992).

Like teachers in other studies (e.g., Nava et al., 1992; Schulz, 1993; Stiggins et al., 1989); Wood et al., 1990), the teachers in this study use both achievement and nonachievement information when formulating grades. However, what is not clear, due to the lack of a well-defined grading system, particularly in American studies, is how this information is reported. Although the majority of the teachers in the present study assumed that a two-grade system was used, the literature was not clear. It appears that standard practice involved using a single grade for achievement and that the task of research, as reported, was to seek whether this one-grade system was contaminated by the teachers' use of nonachievement factors. Although in this study all but one of the teachers agree that a cognitive and a noncognitive dimension should be graded separately (and therefore intend to support the two-grade system), this finding is not apparent in the literature.

As in the studies of Brookhart (1992), Wood et al. (1990), and Nava et al. (1992), all of the teachers include achievement and give it the most weight. There is agreement between the finding in the present study and the findings of Manke and Loyd (1991) and Schulz (1993) that effort can and should be graded. In contrast to Brookhart (1992), the teachers believe that effort should be used pervasively and not only in borderline cases. Additional student characteristics that the teachers in this study include—improvement, potential, attitude, and class participation—were also reported by Griswold and Griswold (1992), Manke and Loyd (1991), Nava et al. (1992), Stiggins et al. (1989), and Wood et al. (1990) in their studies of grading. And as found by Allal (1988), all teachers believe that work habits should be included in a grade.

Process

The finding that 13 of the teachers endorsed a criterion-referenced interpretative framework for grading agrees with the findings of Frary et al. (1992) and Anderson and Bachor (1993), but contrasts with the findings of Bateson (1990) and Wood et al. (1990), who found that teachers used a norm-referenced interpretation, and Allal (1988), who found that teachers used a combination of frameworks. Furthermore, teachers in the present study are generally consistent in assigning grades based on cut-off scores that are compatible with their beliefs, contrary to the finding of Frary et al. (1992) that teachers who endorsed a CRF did not appear to use it in practice.

As found by Allal (1988), Bateson (1990), and Stiggins and Bridgeford (1985), the teachers prefer to use the results from their own "teacher-made" measures to formulate their grades. Like the teachers in Manke and Loyd's (1991) study, the teachers use a variety of measures to determine their achievement grades. Furthermore, the measures that the teachers use most often parallel those reported as most useful by Nava et al. (1992), Wood et al. (1990), and Allal (1988). The use of districtwide measures is not supported by the teachers in this study, in contrast to the finding of Bateson (1990).

Although half of the teachers quantify scores of measures prior to determining grades, subjective methods are also used. This finding is similar to those reported by Allal (1988) and Barnes (1985). In the present study and in both Manke and Loyd's (1991) and Stiggins and Bridgeford's (1985) studies, teachers believe grading to be a highly subjective activity; further, like the teachers in Anderson and Bachor's (1993) and Allal's (1988) studies, the teachers reportedly use intuition.

Context

Although the results of Allal's (1988) study show that contextual factors overruled standard grading practices, the results of the present study reveal that contextual factors play a more encompassing role. Consistent with the teachers in Manke and Loyd's (1991) study, the teachers often make grading decisions based on individual circumstances and their needs. When assigning grades, the teachers either consider the consequences of low-ability students, a practice found by Brookhart (1992), Griswold and Griswold (1992), Manke and Loyd (1991), Wood et al. (1992); or they adjust the requirements for low-ability students (Whitmer, 1983).

The concerns of 17 of the teachers about potential harmful consequences of negative evaluations are similar to those found by Barnes (1985); the modification of practices to avoid these consequences is similar to that found by Brookhart (1992), Mead (1992), and Pilcher-Carlton and Oosterhof (1993). The teachers' hope that grades serve a motivating function is consistent with the findings of Mead (1992). Although the motivational influence of grades is equivocal (teachers gave qualifications in terms of their success rates), the results are somewhat compatible with those reported by Burton (1983).

Similar to the teachers in Wood et al.'s (1990) study, about half of the teachers in the present study reported that they grade academic and nonacademic subjects differently. The teachers in the present study are aware of the conflict between system needs and pupil needs and of the need to balance achievement with the broader social and nonacademic goals of schooling, like the teachers in Barnes' (1985) and Brookhart's (1993) studies. Coupled with this was the repeated reference to the lack of clear criteria for determining grades, a problem identified by Barnes (1985), Griswold and Griswold (1992), and Mehring et al. (1991).

Although the teachers in this study feel that grades are necessary, they have less faith in the ability of grades to reflect student learning accurately, contrary to the

finding of Burton (1983). The teachers' skepticism of the validity of grades is more consistent with the finding of Anderson and Bachor (1993). It is interesting to note that although teachers often referred to the difficulties of grading or indicated a lack of confidence about the accuracy of their grades, this perception contrasts quite markedly with the results of the review of Hoge and C adarci (1989), who found, across the 16 studies they reviewed, that there was a high degree of correlation between the teachers' judgments of their students' academic performance and their students' actual performance on a standardized achievement test.

Conclusions of the Present Study

This study was not meant to discover the right or wrong meanings of grades or the correct way to grade students. It was a discovery of how and why teachers do things in order to find meaning in grading and the grades they provide. In this light, a clearer understanding of teachers' beliefs about grading is gained.

The responses and comments gleaned from this study reveal that the knowledge and expertise of these teachers is extensive. In addition to this knowledge and skill, what they think and do rest on a personal philosophy and the context in which they work; they possess a personal grading plan (Frisbie & Waltman, 1992) or "a philosophical stance" (Terwilliger, 1977, p. 30). They are acutely aware of the practical constraints and realities of the classroom that make recommended practice inappropriate (Stiggins et al., 1989). This in turn influences their personal philosophy and gives them licence to make judgments and assess the appropriateness of certain practices. In short, they employ their own "best practices" (Stiggins et al., 1989, p. 3) as they draw on their experience.

Teachers are deeply aware of how complex grading practices are. Mediating contextual factors are not ignored but taken into account. Grading becomes a 'balancing act' between making adjustments in order to accommodate individual

circumstances and making grades meaningful. These teachers take a child-oriented approach to their work; and, consequently, they often make decisions on an individualized basis. The struggle to maintain a balance between system needs and child needs inevitably favor the child, and therefore they realize that meaningfulness is compromised. Although 10 teachers reported that they do not believe that grades always provide truthful descriptions of student achievement, they strive to make their grading practices meaningful, as conveyed by the following typical statement: "I have to believe that what I am presenting to parents has some validity."

What is somewhat more difficult to interpret is the inherent contradiction between the teachers' beliefs that a formal grading system is necessary for evaluating their students and the repeated reference to grades as limiting and serving as "only one indicator" of student progress. A possible explanation is that although they embrace the concept of 'sorting' and believe that it is necessary, the grade system has not been successfully implemented in their work environment. In their overwhelming support for schoolwide consistency in grading practices, it is clear that teachers, from a logical and practical perspective, are advocating a more common base.

Implications for Practice and Research

Introduction

At the outset of this study a definition of *grade*, which provided a focus for teachers, was given. The key message in that definition concerns the extent to which a grade communicates a common meaning that can be interpreted by various members of a community in a consistent fashion. Whether it is the teachers who assign grades or parents and students who receive them, interpretation hinges on this concept of meaningfulness. As indicated earlier, teachers derive meaning from the functions and processes of their grading practices and the context in which teachers formulate grades and implement their processes.

This section deals with the implications for practice and research from the perspective that meaningfulness is a crucial element and vital to the perpetuation of the grade system. The teachers' responses that revealed that contradictions, complications, and conflict (the 3 Cs) exist in their practices provide a springboard for this discussion. Deficiencies and discrepancies specific to the findings are addressed. The following list of recommendations for practitioners is offered as steps for possible remediation. The order has no affiliation with importance and is set arbitrarily.

A List of Recommendations for Practice

1. In light of the teachers' hope of motivating students with the grades that they give is the reality that using grades to motivate students is at best tenuous. According to Hills (1981), "The motivational aspect of grades can never be primary and must not be allowed to interfere with the communication function." Though it is difficult to erase the impact of grades on motivation, the effects of grades should be monitored, and perhaps teachers should be cognizant of other ways of motivating students. In order to increase the consistency of interpretation, the primary purpose of grades should be the communication of information.

2. Based on the findings in this study, a policy for grading should be developed at the Division 2 (Grades 4-6) level. Though not an exhaustive list, factors to consider in such a policy might include:

(a) What student characteristics should be included in a grade? Teachers should be consistent with what student characteristics are included in the cognitive and noncognitive components that make up the multigrade system. They should keep in mind that too many characteristics in a grade are certain to lead to misinterpretation and that those characteristics that can best or only be evaluated using anecdotal means and/or conferences should not be included in a grade.

(b) On what interpretative framework should grades be based? Further to informing students clearly on what basis they are graded (Manke & Loyd, 1991), teachers should clearly communicate this information to parents, early in the school year. Because of the variation in grading practices across teachers, they are expected to adopt a frame of reference for grading in order to achieve consistency of interpretation. In addition, should they believe that a CRF for grading is most appropriate, they should have a clear understanding of what constitutes that criterion. For example, the criterion can be defined in terms of grade expectations (i.e., the criterion is a sample of work designated as average for a given level) or in terms of specific criteria (the criterion is mastery of a set of listed objectives). Furthermore, the criterion may differ across subjects.

(c) What procedures should be used when aggregating the scores from various measures prior to assigning a grade? The scores from each of the measures should be in the same score scale and weighted to reflect the intended emphasis given in the program of studies (*Principles*, p. 11). Numerical scores may be averaged and converted to the letter symbol which best represents the amount of learning which has taken place.

(d) How should students of differing ability be graded? Although more troublesome to resolve, the issue of grading according to a student's ability has implications for practice. One problem concerns the issue of grading students of 'regular' classes who differ in ability. It is recommended that expectations of students should be a function of their beginning state of knowledge and skill. A second problem is the grading of borderline students, particularly if a low achiever is subject to potential harmful or negative consequences. Because there may be a need to individualize grading practices, proper documentation is required if the information is to be useful. For example, the need for a receiving teacher to receive accurate

information may need to be balanced with confidentiality of information. Directives for accommodating individual cases should be in place.

(e) How should academic and nonacademic subjects be graded? To increase the meaningfulness of grades, interpretative frameworks, procedures for determining grades, and the identification of appropriate symbols to use should be made explicit for the grading of both academic and nonacademic subjects. Teachers should be aware that prior to the valid interpretation of any grade, the assessment methods which serve as components of grades must be relevant and representative of the performance assessed (*Principles*, p. 12).

(f) What measures should make up the components of grades? The reported preference for teacher-made assessments implies that common measures may need to be developed if consistency of meaning of a grade (validity) is to be realized. In addition, scoring guides may need to be developed if grading practices are to be consistent (reliability). However, this factor demands a comprehensiveness that may be beyond the practicality or manageability of a particular school to implement, so a workable balance should be set. Teachers should also exercise caution when using daily assignments for grades. They should be mindful that assignments are useful for grades only when they reflect the amount of learning that has taken place and that basing grades on completion of work is not appropriate. By the same token, it is recommended that homework remain an instructional device and be used as practice in preparation for assessments.

3. In order to increase teacher confidence and expertise, workshops should be conducted to assist teachers with deficiencies. As indicated in the data, teachers appear to need assistance regarding the choice of framework to employ when grading and how to combine scores and determine cut-off points for their grades. There may also be a need to determine how effort should be used in grading and how teachers can 'manage' subjectivity when involved in the grading process.

Implications for Research

1. This study or parts of it could be replicated at other levels of schooling or at different locations. For example, certain questions could be adapted to the primary level. A replication at a rural setting could be used for a comparison to the present findings.

2. This study should be replicated using a method designed to capture a larger representation of teachers, particularly where inconsistencies exist. Using a survey or questionnaire, research might be conducted in one or more of the following areas:

(a) Should students in regular classes be graded according to ability or to a common standard?

(b) How should effort be defined? Is effort best indicated by intense performance or sustained performance (Griswold & Griswold, 1992)? Should an effort grade include work habits or other student characteristics?

(c) Should grades of academic and nonacademic subjects be graded in the same ways?

(d) Should the broader outcomes of schooling be graded? In terms of social skills, independent working habits, or attitude, for example, are the interests of children best served by a comparison of these attributes with standards/norms or by identifying those attributes that are unique to students and thereby individualizing evaluation?

3. In order to provide direction for policy making, research may be needed to determine the current status of grading policies. Information of the following types would be appropriate:

(a) What grading policies exist (if any) and at what administrative level(s)?

(b) How are their contents communicated to teachers? Is there a written policy?

(c) What are the main purposes of grades, and for what will they be used?

(d) Does the policy give teachers specific information for the implementation of the grading plan? For example, is there a frame of reference for grading in place? Are criteria for grades specified? Are they weighted in cases where multiple criteria are used? Are all subjects graded in a consistent way?

(e) How are grades for borderline or other individual cases treated?

4. There is concern about the misinterpretation of grades on report cards. Research should be conducted to determine how various audiences interpret grades. The validity of report cards could be determined in a study designed to provide information on various features included in report cards.

References

Allal, L. K. (1988). Quantitative and qualitative components of teachers' evaluation strategies. *Teaching and Teacher Education, 4*(1), 41-51.

Anderson, J. O., & Bachor, D. G. (1993). *Assessment practices in the elementary classroom: Perspectives of the stakeholders.* Paper presented at the annual meeting of the NCME, Atlanta, GA.

Austin, S., & McCann, R. (1992). *Here's another arbitrary grade for your collection: A statewide study of grading practices.* ED 343944.

Barnes, S. (1985). A study of classroom pupil evaluation: The missing link in teacher education. *Journal of Teacher Education 38*(4), 46-49.

Bateson, D. J. (1990). Measurement and evaluation practices of British Columbia science teachers. *Alberta Journal of Educational Research 36*(1), 45-51.

Brookhart, S. M. (1992). *Teachers' grading practices: Meaning and values.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Brookhart, S. M. (1993). *Grading and classroom management: What does it mean to earn a grade?* Paper presented at the annual meeting of the AERA, Atlanta, GA.

Burton, F. (1983). *A study of the letter grade system and its effect on the curriculum.* ED 238143.

Constas, M. A. (1992). Qualitative analysis as a public event: The documentation of category development procedures. *American Educational Research Journal, 29*(2), 253-266.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*(4), 438-481.

Dorr-Bremme, D. W. (1983). *Assessing students: Teachers routine practices and reasoning.* ED 242742.

Ebel, R. L. (1974). Shall we get rid of grades? *Measurement in Education, 5*(4), 1-4.

Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Frary, R. B., Cross, L. H., & Weber, L. J. (1992). *Testing and grading practices and opinions in the nineties: 1890s or 1990s?* ED 351357.

Friedman, S. J., & Frisbie, D. A. (1993). *The validity of report cards as indicators of student performance.* Paper presented at the Annual Meeting of NCME, Atlanta, GA.

Friedman, S. J., & Manley, M. (1991). *Grading practices in the secondary school: Perceptions of stakeholders.* Paper presented at the Annual Meeting of NCME, Chicago, IL.

Frisbie, D. A., Waltman, K. K. (1992). Developing a personal grading plan. *Educational Measurement: Issues and Practice, 11*(10), 35-42.

Griswold, P. A. (1988). *Teacher attitudes towards testing.* ED 305363.

Griswold, P. A., & Griswold, M. M. (1992). *The grading contingency: Graders' beliefs and expectations, and the assessment ingredients.* Paper presented to the 1992 Annual Meeting of the American Research Association, San Francisco, CA.

Gronlund, N. E., & Linn, R. L. (1990). *Measurement and evaluation* (6th ed.). New York: Macmillan.

Gullickson, A. R. (1984). Teacher perspectives of their instructional use of tests. *Journal of Educational Research, 77*(4), 244-248.

Gullickson, A. R. (1985). Student evaluation techniques and their relationship to grade and curriculum. *Journal of Educational Research, 79*(2), 96-100.

Gullickson, A. R. (1986). Teacher education and teacher-perceived needs in educational measurement and evaluation. *Journal of Educational Measurement, 23*(4), 347-354.

Hargis, C. H. (1990). *Grades and grading practices.* Springfield, IL: Thomas.

Hills, J. R. (1981). *Measurement and evaluation in the classroom* (2nd ed.). Columbus, OH: Merrill.

Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research, 59*(3), 297-313.

Hopkins, K. D., Stanley, J. C., & Hopkins, B. R. (1990). *Educational and psychological measurement and evaluation* (7th ed.). Englewood Cliffs, NJ: Prentice-Hall.

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

Leiter, J., & Brown, J. S. (1983). *Sources of elementary school grading.* North Carolina. ED 236135.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry.* Beverly Hills, CA: Sage.

Linn, R. L. (1990). Essentials of student assessment: From accountability to instructional aid. *Teachers College Record, 91*(3), 422-436.

Manke, M. P., & Loyd, B. (1991). *A study of teachers' understanding of their grading practices.* Paper presented at the meetings of the National Conference on Measurement in Education, Chicago, IL.

Marso, R. N., & Pigge, F. L. (1992). *A summary of published research: Classroom teachers' knowledge and skills related to the development and use of teacher-made tests.* Paper presented at the Annual Meeting of the AERA, San Francisco, CA.

McMillan, J. H., & Schumacher, S. (1989). *Research in education: A conceptual introduction* (2nd ed.). USA: HarperCollins.

Mead, J. (1992). *Teachers' evaluations of student work.* ED 348357.

Mehrens, W. A., & Lehmann, I. J. (1991). *Measurement and evaluation in education and psychology* (4th ed.). Toronto, ON: Holt, Rinehart & Winston.

Mehring, T., Parks, C., Walter, K., & Banikowski, A. (1991). Report cards: What do they mean during the elementary school years? *Reading Improvement, 28*(3), 162-168.

Miles, M. B., & Huberman, A. M. (1984). *Qualitative data analysis: A sourcebook of new methods.* Beverly Hills, CA: Sage.

Nava, F. E., Josefa, G., & Loyd, B. H. (1992). *An investigation of achievement and non-achievement criteria in elementary and secondary school grading.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Oosterhof, A. C. (1990). *Classroom applications of educational measurement.* Columbus, OH: Merrill.

Pilcher-Carlton, J., & Oosterhof, A. C. (1993). *A case study analysis of parents', teachers', and students' perceptions of the meaning of grades: Identification of discrepancies, their consequences, and obstacles to their resolution.* Paper presented at the Annual Meeting of the AERA, Atlanta, GA.

Popham, W. J. (1990). *Modern educational measurement: A practitioner's perspective* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Principles for fair student assessment practices for education in Canada. (1993). Edmonton, AB: Joint Advisory Committee, Centre for Research in Applied Measurement and Evaluation, University of Alberta.

Scates, D. E. (1943). Differences between measurement criteria of pure scientists and of classroom teachers. *Journal of Educational Research, 37,* 1-13.

Schulz, H. (1993). *Grading and reporting practices in Newfoundland and Labrador schools.* Paper presented at the 1993 Annual Meeting of AERA, Atlanta, GA.

Simon, S. B., & Bellanca, J. A. (Eds.). (1976). *Degrading the grading myths: A primer of alternatives to grades and marks.* Washington, DC: Association for Supervision and Curriculum Development.

Stiggins, R. J. (1988). Revitalizing classroom assessment: The highest instructional priority. *Phi Delta Kappan, 69*(5), 363-368.

Stiggins, R. J. (1991). *A practical guide for developing sound grading practices.* Portland, OR: Northwest Regional Educational Laboratory, Center for Classroom Assessment.

Stiggins, R. J., & Bridgeford, N. J. (1985). The ecology of classroom assessment. *Journal of Educational Measurement, 22,* 271-287.

Stiggins, R. J., & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment.* Albany, NY: State University of New York Press.

Stiggins, R., Frisbie, D., & Griswold, P. (1989). Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and Practice, 8*(2) 5-14.

Strauss, A. L. (1987). *Qualitative analysis for social scientists.* New York: Cambridge University Press.

Terwilliger, J. S. (1971). *Assigning grades to students.* Glenview, IL: Scott, Foresman.

Terwilliger, J. S. (1977). Assigning grades: Philosophical issues and practical recommendations. *Journal of Research and Development in Education, 10*(3), 21-39.

Terwilliger, J. S. (1989). Classroom standard setting and grading practices. *Educational Measurement: Issues and Practice, 8*(2), 15-19.

Thiessen, D., & Moorhead, R. (1985). *More than marks: What teachers say about student evaluation.* Toronto: Ontario Public Schools Teachers' Federation.

Thorndike, R. L., & Hagen, E. (1977). *Measurement and evaluation in psychology and evaluation* (4th ed.). New York: Wiley.

Waltman, K. K., & Frisbie, D. A. (1993). *Parents' understanding of their children's report card grades.* Paper presented at the Annual Meeting of the AERA, Atlanta, GA.

Whitmer, S. P. (1983). *A descriptive multimethod study of teacher judgment during the marking process.* ED 234052.

Wiggins, G. (1988). Rational numbers; Toward grading and scoring that help rather than harm learning. *American Educator, 12*(4).

Wood, P., Benner, T., & Wood, J. (1990). *Grading and evaluation practices and policies of school teachers.* Paper presented at a joint session of the Annual Meetings of the National Council on Measurement in Education and the American Educational Research Association, Boston, MA.

Appendix A

Documents: Ethics Approval, Permission to Conduct Research, and Request for Teachers' Participation

March 16, 1994

From: Department of Educational Psychology
Research and Ethics Committee

The Research and Ethics Committee of the Department of Educational Psychology has reviewed the attached proposal and finds it acceptable with respect to ethical matters.

Applicants: Dr.W. Todd Rogers on behalf of Elizabeth A. Kushniruk (graduate student)

Title: Teachers' Beliefs and Philosophies of their Grading Practices

Participating Agency(ies):

Recommended Changes:

P M Julicah
Chairman or Designate, Research and Ethics Committee

March 17/94
Date

EDMONTON PUBLIC SCHOOLS

March 24, 1994 (File #052.94)

Mr. Randy Wimmer, Director
Practicum Placement
Division of Field Services
University of Alberta
Edmonton, Alberta
T6G 2G5

Dear Mr. Wimmer:

Re: Research Request: Teachers' Beliefs and Philosophies of Grading Practices: Kushniruk (Rogers)

The above research request has been approved on a permissive basis following examination by our department. The approval is subject to the following conditions.

1. Teacher participation in the study to be voluntary;

2. Teachers are free to withdraw at any time;

3. The results of the study will be provided to the teacher;

4. Anonymity of the teachers and the confidentiality of information obtained is assured; and

5. The researcher provides a copy of the results to this office.

Elizabeth Kushniruk should now proceed with inviting those teachers in the sample we have selected to participate in the study.

We suggest she include a copy of this letter of authorization in the envelope sent to the teachers.

I wish you success with the project and look forward to receiving a copy of the results. (PLEASE QUOTE FILE # ABOVE)

Yours sincerely,

Simon van der Valk
Supervisor Monitoring and
Student Information

SVV/LH/clt

cc: Elizabeth Kushniruk

March 23, 1994

Dear Fellow Teacher,

My name is Elizabeth Kushniruk and at present I am on leave from Edmonton Public Schools completing my master's degree at the University of Alberta. I am writing this letter to ask you for your participation in a study I am conducting regarding teachers' beliefs about grading.

You have been randomly selected by personnel at Edmonton Public Schools to assist me with my study. They have agreed to distribute this information on my behalf; however, your identity has not been reveaied to me at this time. Please find enclosed a letter indicating that I have received permission to conduct my research.

By having the opportunity to interview you, I hope to be able to gain insight into your thoughts about what meaning grades have for you and how you think as you grade your students' work. In light of the current focus on assessment and evaluation, there has been little research done on what teachers believe about grades and the grading process. I believe that this type of research is timely for us, and to my knowledge it has not been done within our district. This study will give me a chance to use your valued input on this important topic.

Your participation in the study will be in the form of an interview, which will be audiotaped and replayed in order to accurately describe and document your responses. I will be willing to meet with you at a convenient time and location. During the interview I will ask you specific questions and give you the opportunity to reply freely as well. It is important that you describe your true beliefs about how grading practices should be conducted. There are no 'right' or 'wrong' answers. I am interested in what you, as a fellow teacher, hold to be your values and beliefs about grading.

All your responses will be treated with strict confidentiality, and you may assume a fictitious name in the proceedings. Should you decide to withdraw from the study at any time, you may do so without prejudice. When I have completed my study I will be willing to share my findings with you.

I am looking forward to meeting you as a participant in this study. Please call me at my home (456-7061) to arrange for a suitable time and place for our interview. If you require more information or clarification, I would be happy to comply. In order to collect data prior to the 'busy' time of the school year, I am asking that you contact me by April 12th.

Yours truly,

Elizabeth Kushniruk

Appendix B

Interview Schedule

Interview Schedule

Name _____ Date _____

Grade Level _____ M or F ___ Years of Experience ____

Measurement course(s) taken (if any) _____

1. What does a grade or the process of grading mean to you? (Open response)

   Prompt: What are you doing when you grade your students?

   Prompt: What meaning do you wish to convey when you assign grades to your students?

   [Show my definition of grade/grading to the participant and explain that this meaning will be the frame of reference in all subsequent questions. Explain how summative evaluation directly reflects the grade assigned and how formative evaluation should not contribute to grading decisions.]

2. (a) Grades are assigned and reported in order to communicate information about student achievement or progress. Who should be the primary audiences to receive the information that grades give? (Open response)

   (b) Here is a list of possible audiences. What do you believe they should know or want to know?

   [Hand a card to the participant for check-off.]

   the student _____
   the parent(s) _____
   the administration in schools _____
   the district or province _____
   future institutions or employers _____
   others _____

3. (a) Do you hope, in some way, to influence students with your grades?

   ____ Yes If yes, how?

   ____ No If no, why not?

   (b) [If motivation or equivalent is not mentioned, then ask:]

   Do you feel it is likely that students are motivated by grades?

   ____ Yes If yes, how?

   ____ No If no, why not?

(c) [If motivation is mentioned above and nothing else, then ask:]
Do you believe students may be influenced by grades in additional ways? If so, how?

(d) Do you have any concerns about any of these influences?

____ Yes If so, which ones, and why?
____ No

(e) It has been suggested that the consequences of giving a particular grade need to be taken into consideration. Do you feel that it is more important that grades reflect the current status of student achievement than that they be used as guides for students towards some future course of action?

____ Current
____ Future

Why do you feel this way?

4. (a) Of the following student characteristics, which ones do you believe should be assessed and factored into the computation of a grade?

[Hand the participant a card for check-off.]

____ achievement
____ effort
____ aptitude or potential to learn
____ improvement (from previous grading period)
____ work habits
____ attitude/behavior
____ class participation
____ citizenship/co-operativeness

Which ones are most important? Are they equally important?

(b) [Show my definition of achievement and nonachievement variables to the participant and say:]

It has been suggested that grades should reflect only achievement and that other student characteristics should be reported separately. How does this compare with your own philosophy?

(c) What do you believe to be true about the actual practice among teachers?

(d) Are there situations where you would 'bend the rule' and consider other factors when assigning a final grade? (Open response)

If so, what are these situations?

Do you believe that this is fair?

5. (a) Of the three frameworks used for grading [refer to my earlier definition], which do you believe to be the most appropriate at your grade level?

____ NRT
____ CRT
____ Ipsative

Please give a reason for your choice.

[If ipsative is chosen, ask how comparison to oneself should be entered into a grading decision.]

(b) During the reporting period you keep a record of a number of different student performances. Do you use a student's total scores prior to assigning a grade?

____ Yes
____ No

If no, then how do you combine the various different results of a student's performance to make a decision about a grade?

(c) Suppose you had a class set of total scores, say out of 100. How should cut-off scores be established for mapping these totals on a grade scale?

[Hand a card to the participant showing hypothetical set of class scores.]

____ Preset distribution
____ Preset standard
____ Empirically derived

Why do you feel that this is the best procedure to use?

(d) Please rank the usefulness of the following symbols to reflect student performance.

[Hand a card to the participant for check-off.]

____ At, above, or below grade level
____ Percentage; i.e., 85%, 61%
____ Excellent, good, fair, poor
____ Pass/fail, satisfactory/unsatisfactory,
____ A, B, C, D, F
____ Others

[If growth was not suggested earlier as entering into the grading decision, say:]

(e) It has been suggested that the 'growth' factor be included in the reporting of student progress. How should student growth be reported?

6. (a) Should different 'ability' groups be graded according to different criteria?

____ Yes
____ No

Can you give your reasons?

(b) Should programs with different objectives be graded differently?

[Probe:] For example, teachers may feel that math and music have quite different objectives and desired outcomes.

____ Yes
____ No

(c) Do you feel that grading practices should be consistent across the four core academic subjects?

____ Yes
____ No

(d) Should they be consistent within a particular school?

____ Yes
____ No

(e) Some districts employ common grading policies for all teachers to use, whereas others do not. What do you feel is right for our district regarding grading practices?

7. Do you believe that grades provide truthful and consistent descriptions of student achievement, such that you can make informed decisions?

____ Yes
____ No
____ Don't really know

[Probe:] Is it possible to evaluate the 'worth' of a student without the assignment of a grade?

8. (a) Are you able to grade students according to your beliefs, or do you have to follow a school policy which is contrary to your beliefs?

(b) Can you identify one or two things that are most troublesome for you during the grading process? (Open response)

(c) Do you feel that guidance is needed at the school/district level for developing staff expertise and confidence in grading practices?

____ Yes
____ No

If yes, what suggestions would you make?

9. I have one final question which is related to grading. In what ways should achievement be measured so that it can be of use in establishing grades?

[Hand a card to the participant for check-off.]

____ assignments and class projects
____ oral questions
____ paper and pencil tests, quizzes
____ observations and judgments (rating scales or checklists)
____ intuition and feelings
____ homework
____ standardized tests; i.e., CGAT, CTBS, GATES
____ district-wide tests
____ others

[If standardized tests or district tests were omitted, then ask:] Should they be reported separately?

[Thank the participant for all of her/his important contributions.]

Appendix C

Study Description/Informed Consent Form

Study Description/Informed Consent Form

My name is Elizabeth Kushniruk, and I am a master's student in the basic program of the Educational Psychology Department at the University of Alberta. My thesis involves an investigation into the beliefs and philosophies that teachers have about grading practices.

You have been selected in a random sample to assist me with my study. By having the opportunity to interview you, I hope to be able to gain insight into your thoughts about what meaning grades have for you and how you think as you grade your students' work. Although the literature is rapidly accumulating on various assessment practices and evaluative techniques, it appears to be lacking in the specific area of grading and reporting. Furthermore, although several texts and manuals describing ways of grading have been written, there lacks specific information of what teachers believe to be important and relevant as they go about the process of grading students' work.

Your participation in the study will be in the form of an interview, which will be audiotaped and replayed in order to describe and/or document your responses accurately. The time and location of the interview will be mutually decided.

During the interview I will ask you specific questions and give you the opportunity to reply freely. It is important that you describe your beliefs truthfully about how grading practices should be conducted. There are no right or wrong answers. I am interested in what you, as a practicing teacher, hold to be your values and beliefs about grading.

All your responses will be treated with strict confidentiality, and you may assume a ficticious name in the proceedings. Should you decide to withdraw from the study at any time, you may do so without prejudice. When I have completed my study I will be willing to share my findings with you.

Should you have any further questions please do not hesitate to call me at my home (456-7061).

I hereby give my consent to participate in the study as described above.

Name: _____

Date: _____