THE EFFECTS OF SELECTED STUDENT ERRORS ON THE CONSISTENCY OF TEACHER GRADING IN MATHEMATICS

PATRICIA JEAN MAXWELL

THE EFFECTS OF SELECTED STUDENT ERRORS ON THE CONSISTENCY OF TEACHER GRADING IN MATHEMATICS

by

© Patricia Jean Maxwell, B.Sc., B.Ed.

A Thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Education

Department of Curriculum and Instruction
Memorial University of Newfoundland

St. John's Newfoundland

August 1982

Abstract

The purpose of this study was to examine the consistency of teachers in the identification, classification, and grading of student solutions to grade 10 Algebra word problems. Also investigated were the effects on assigned grades, of error location, error type, and whether or not a correct numerical solution was obtained.

The sample group was randomly selected from the population of all teachers of grade 10 Academic Mathematics in the Province of Newfoundland for the school year 1981-82. Of the 100 schools selected, 69 responded before the deadline amounting to a total of 90 teachers.

The results indicated that teachers were very consistent in the location of student errors. In fact, the errors were successfully located 99.4% of the time. They were slightly less consistent in the classification of these errors. Teachers used both general and specific classification terminologies to describe the student's errors and in a few instances simply referred to the error as reflecting careless or sloppy work. The grading of these errors showed much greater variability. Teacher's grades were most consistent when errors were made near the beginning of solutions and least consistent when errors were made near the end. The standard deviations of the scores of individual items ranged from a low of 1.07 to a high of 1.97. Grading

practices were found to vary also in the level of severity individuals maintained. The amount the grades were distributed throughout the grading scale and the relative judgments of importance that were attributed to the various error types.

It was also found that teachers graded property errors more severely than computational errors. Furthermore, errors made near the beginning of solutions appeared to be graded more severely than equivalent errors made near the end of a solution. No conclusive relationship was apparent between the grades teachers assign and whether a correct or incorrect numerical solution was obtained.

Acknowledgements

I would like to thank all the teachers who participated in this study for their cooperation. A special thanks is extended to the teachers from those schools that participated in the pilot study. These teachers were especially helpful and obliging.

Sincere thanks is also extended to my husband, Rob, for his help and support throughout this project.

Special appreciation is extended to my supervisor, Dr. Lionel Mendoza, for his help, interest, and assistance, from the beginning to completion of this study.

Table of Contents

Page

List of Tables

List of Figures

CHAPTER I

THE PROBLEM

Introduction

The importance of evaluation and measurement in education should not be underestimated. In particular, as Ebel (1972) suggests, consistent and objective measurement of student performance has long been a concern of educators. It is important, however, before any discussion of evaluation to differentiate between measurement and evaluation. Many definitions of student evaluation have been put forward. Gronlund (1971) defined evaluation as, "a systematic process of determining the extent to which educational objectives are achieved by pupils." (pp. 7-8) Similarly, Ebel (1972) defined evaluation as, "a qualitative judgment of how good or how satisfactory the students performance has been." (p. 326) The judgment of how satisfactory student performance has been is based in whole or in part on measurements of achievement, the other components being the value judgments made by the teacher. Ebel (1972) defined measurement as, "a quantitative description of how much a student has achieved." (p. 326) Although writers such as Gronlund (1971) and Ebel (1972) emphasize the use of several means to assess student achievement, much of the information used in determining how much a student has achieved at the high school level is obtained by the use of achievement tests.

The measurement of student performance is one of the most difficult tasks that a teacher undertakes. Many writers, such as Ebel (1965), Gronlund (1971), and Lien (1976), began their discussion of marks and marking systems by using the introductory phrase: 'The problem of marking'. This phrase is probably indicative of how many teachers view marking. Davis (1964) pointed out that, "probably the most disliked aspect of a teacher's job is marking." (p. 287)

However, as Ebel (1972) indicated, many teachers are unaware of the difficulties associated with marking, the fallibility of their judgments, and the personal biases that may be reflected in their marks. Hadley (1954) included the following factors as components of the marks teachers assign to their pupils: "actual attainment, the teacher-pupil relationship, deportment, sex, promptness and attendance, personal appearance, obedience, effort, and attitude." (p. 305)

Despite the problems inherent in it, marking has played and will continue to play an important role in education. School grades are used for a variety of purposes. These purposes can be viewed as both administrative and instructional. Writers such as Adams and Torgerson (1964), Ebel (1972), and Lien (1976) have discussed these uses and are in general agreement. The first of these uses is administrative, where grades are used for purposes of transfer, promotion, graduation, honor roles, etc. Secondly, they are used for motivational purposes, where grades are considered to provide incentive for greater effort on the part of the student. Even though motivation is considered as a use of grades by some, Terwilliger (1971) pointed out that the motivational function is more a consequence of evaluation than a purpose, per se. Thirdly, grades are used for the purpose of guidance. Grades are used to help in the identification of the strengths and weaknesses of the student so that realistic planning can take place for the student's educational future. Finally, grades are used for informational purposes, that is, the transmission of information to students and parents on the student's progress.

As indicated by Ebel (1972), progressive schools and colleges are constantly experimenting with new systems of grading and sometimes with not grading at all. However, as yet no viable alternatives have been found. Suydam (1974) indicated that tests and, therefore, the necessity of grading is going to be a part of our educational environment for a long time to come. The grading of tests provides the most feasible way of finding out, relatively quickly, how well each child has learned certain content. In further support of the continued necessity for grading, Thorndike and Hagan (1977) stated:

> Marks and marking have been deeply embedded in the educational culture. They have become the basis, in whole or in part for a wide range of actions and decisions within a given educational structure and in the relations of the educational system to the outside world. (p. 588)

They went on to say that like other aspects of a culture that have become deeply ingrained, marks as well as marking procedures, are often taken for granted. Since marks are used for such a variety of purposes and no better means of student evaluation have yet been put forward than measurement through testing, a study of teacher consistency of marking student test responses can be considered as a worthwhile direction for research.

Rationale

From a search of the literature it was found that little research has been conducted in the area of student evaluation in mathematics, per se. The work that has been published has been largely concerned with the design and improvement of the teacher-made test, such as the studies of Taylor (1966), Coppedge and Hanna (1971), and Majors and Michael (1975).

Teacher grading of these tests has received relatively little consideration. Ebel (1972) observed,

> From some points of view it (marking) is even more complex and difficult than the problem of building a good test and using it properly. (p. 307)

Much of the research which has been conducted in grading work in mathematics has focused on teacher bias in grading. This research (Carter, 1952 and Duval, 1980) has largely been concerned with whether the sex of the student or the sex of the teacher had any effect on the grades teachers assign.

In the area of English composition grading, there have been a few studies which have more specifically attempted to identify the effects of selected errors on the grades assigned by teachers. Scannell and Marshall (1965) and Marshall (1966) both undertook research to determine the effects of errors in grammar, punctuation, and spelling on the grades teachers assigned to students' work.

Thorndike (1971) pointed out that it is very difficult to achieve consistency in grading essay responses to questions and gave three 'inescapable' conclusions related to this difficulty. They were:

> (a) Different raters tend to assign different grades to the same paper...
> (b) A single rater tends to assign different grades to the same paper on different occasions.
> (c) The differences tend to increase as the essay questions permit greater freedom of response. (p. 277)

Thorndike also identified several reasons for these conclusions:

> First, raters may differ in their severity. One characteristically may assign relatively high grades while another may tend to assign generally low grades... (Secondly) they differ in the extent to which they distribute grades throughout the score scale. Some tend to distribute scores closely around their average; others spread scores much more widely.
>
> ...Finally, raters may differ in the relative value they assign to different papers. A paper judged by one rater to warrant a high grade may be judged inferior by another. (p. 277)

Although Thorndike's references to inconsistency of grading refer specifically to essay questions, the ideas expressed do have relevance with other forms of testing as well. At least this would appear to be true if that subject is evaluated by the use of an instrument which requires decisions of relative weighting of aspects of answers or the development of personal marking schemes on the part of the grader. When content is tested by very objective instruments such as, multiple-choice, true-false, or matching formats grading consistency would seem to be relatively easy to maintain.

A major implication of the variations that can exist in teacher grading is that a student may be failing with one teacher in one school, while the same student could conceivably see a measure of success with another teacher or in another school. In the Shared Evaluation Handbook (1980) developed through the Department of Education for the Province of Newfoundland and Labrador, it was reported that 'standards' of marking do vary from school to school. It was due to the recognition of this inconsistency that a modification plan, referred to as the "D factor", was developed. This modification plan was designed to bring individual teacher's marking standards into line with the provincial average. In this way students who possess similar ability are more likely to receive similar grades independent of the school which they attend. The "D factor" is applied to help control for teacher inconsistencies in grading at the grade 11 level. However, it must be recognized that these variations exist, probably to the same extent, at other grade levels as well.

Starch and Elliott (1912, 1913) studied the reliability of teacher grading in mathematics, English and history. They found that a great deal of variation existed at all three subject areas. It was found that grading in mathematics showed even greater variation than English or history. These researchers indicated that such results ran contrary to the popular belief of the time which viewed mathematics as an exact science that could be graded with 'mathematical precision'.

From the work of researchers such as Starch and Elliott (1912, 1913), Carter (1982), and Duval (1980) it is evident the variation and inconsistency in teacher grading of mathematics exists. Therefore, it would be worthwhile to ascertain the extent of these inconsistencies, by identifying the areas of greatest and least variability, and more specifically, to identify some of the factors which play a role in the determination of the actual grade a teacher assigns to a student's answer. Through the identification of these factors a better understanding of the things which contribute to a possible lack of consistency in mathematics grading can be established.

Purpose of the Study

In this study the consistency of teacher grading of specific word problems containing various error types was examined. Also investigated were the effects, on assigned grades, of error location, error type, and whether or not a correct numerical solution was obtained.

More specifically answers to the following questions were sought.

(i) Are teachers as a group consistent in the identification of the errors made by students?

(ii) Are teachers as a group consistent in their grading of student responses?

(iii) Does the type of error make a difference to teachers grading of student responses?

(iv) Does the location of the error make a difference to the teachers grading of student responses?

(v) Does the correct answer make a difference to teachers grading of student responses?

Definition of Terms

For the purpose of this study, the following terms required definition:

Word Problems: Problems which require the analysis of a verbal statement in order to set up and solve an equation and to interpret the solution as it relates to the original problem.

Scoring:	The actual process of correcting assignments, tests, projects, etc.
Grading/Marking:	The assigning of a grade to tests or assignments that are being evaluated.
Academic Mathematics:	The core Mathematics program designed for the majority of students in the Province of Newfoundland (Division of Instruction, Department of Education, Government of Newfoundland and Labrador, 1981).

Delimitations

There were several delimitations evident in this study. The word problem is in itself just one aspect of Algebra and the grading practices used there may not be representative of the practices applied to other content areas. Also, the study was delimited to just one grade level. The same error made by a grade 9 or 11 student might have been graded differently since the teachers of these grade levels may expect a completely different standard of work. Finally, it cannot be assumed that teachers of other grade 10 mathematics programs would show the same leniency or rigor in grading and thus produce the same grading distributions.

Outline of Report

A review of the related literature is presented in Chapter II of this report. It covers three main areas; student evaluation in general, research carried out in other subject areas closely relating to the purpose of this study, and research specifically related to student evaluation in Mathematics. The sampling procedure, the development, piloting and distribution of the instrument, the limitations of the study, and the methods used to analyze the data are described in Chapter III. In Chapter IV the results of the study are reported. In Chapter V these results and their implications are discussed and areas for further study are recommended.

CHAPTER II

REVIEW OF RELATED LITERATURE

In this chapter the literature related to student evaluation in mathematics is reviewed, focusing, where possible, on the grading of students' work. The chapter is organized into three main sections. In the first section, literature related to the importance and uses of student evaluation in education is discussed. In the second section, research on evaluation conducted in subject areas other than mathematics but related to the purpose of this study is reviewed. Finally, research relating to student evaluation in mathematics is reviewed. This section concentrates on three main areas. They are the design and improvement of teacher-made tests, the grading practices of teachers as they relate to homework assignments in mathematics, and the bias of teachers in their grading of student's work.

Student Evaluation

> Evaluation is a continuing, integral aspect of mathematics teaching, concerned with the improvement of instruction. Evaluation ascertains whether the teacher is teaching what he thinks he is teaching and the learner is learning what the teacher thinks the learner is learning. Evaluation is qualitative as well as quantitative: it involves appraisal as well as measurement, for it includes the stage of making value judgments. (Suydam, 1974, p. 5)

The discussion of evaluation presented in this section concentrated mainly on that component of evaluation commonly described as measurement. While writers such as Gronlund (1971), Ebel (1972), and Suydam (1974) all emphasized qualitative judgments as a significant component of the total picture of student evaluation, in this study judgment of student performance on the quantitative level was the emphasis; therefore, the literature reviewed is centered primarily on this aspect.

There are two major components which are essential for effective measurement of student performance on the quantitative level. They are the design and implementation of a measuring instrument and the planning for and implementation of a grading/marking system. These can be described as the tools of measurement. As Ebel (1972) indicated, marking can be an even more difficult task than the construction of a good test.

The general uses of measurement in education, namely administrative, motivational, guidance, and informational, were discussed in some detail in Chapter I. In addition Thorndike and Hagan (1977) suggested that grades also play a role in building the students self-image. Students use grades as a means of judging what is important for them to learn and identifying what they are good at doing. Thorndike and Hagan (1977) indicated that:

> what the teacher assigns as learning tasks, and more especially what he or she corrects, grades, and returns to the pupils defines for them what is considered important in school. (p. 590)

This notion was supported by the findings of White and Boehm (1967). These researchers found that elementary school students tended to believe spelling and arithmetic were the most important things they studied in school because those were the subjects that the teacher most often graded.

Somewhat in contrast with Thorndike and Hagan's (1977) suggestion that grades play a role in building the student's self-image writers such as Gaier (1966), Lien (1976), and Hill (1976) warned of some possible negative side-effects of grades. For example, they can become so important that learning is judged according to the numerical value that can be placed on it, rather than on the quality of learning that takes place. Grades, too often, become ends in themselves. Ahman and Glock (1981) indicated that the negative by-products of marking result in emotional problems for students. Gaier (1966) suggested that when students do not do well repeatedly, the motivation to try slowly erodes. Other negative effects of grades occur when teachers use marks as rewards or prizes or when pressures to earn high marks result in student cheating.

The classroom teacher is often unaware of the broader, more global uses and possible misuses of evaluation and the potential impact on the student. The teacher's main aim in evaluation is to determine whether the student has acquired certain content. Terwilliger (1971) identified three distinct reasons, at the classroom level, the teacher has for evaluating the achievement of students. They are:

> 1. To determine whether a student has acquired specific skills and knowledge considered essential by the teacher.
>
> 2. To diagnose strengths and weaknesses of a student as revealed by the patterns of achievement demonstrated.
>
> 3. To determine the achievement of a student with reference to others who have received comparable instruction. (p. 6).

These reasons relate to the instructional decisions required of the teacher as well as the providing of information to the student and the teacher which can be used for purposes such as making course selections, decisions relating to the need for extra help, or extra study.

In the previous discussion it was suggested that school grades are used for a variety of purposes, both in terms of the value placed on them by society and the needs they satisfy at the classroom level. The school is an agent of society and grades provide a means of describing for society what individuals have accomplished within the school environment. A concise description of why grades are so necessary to schools was given by Terwilliger (1971). He wrote:

> The purpose of grading systems is to provide a systematic and formal procedure for transmitting value judgments made by the teacher to the student and to others

> most directly concerned with his development and welfare. These value judgments provide a basis for making important decisions which are faced in the normal course of an individual's development in our society. (p. 7)

Research on Evaluation in Subject Areas Other Than Mathematics

In this section the research carried out in subject areas other than Mathematics is surveyed, focusing specifically on the reliability of teacher grading and the factors which influence the grades teachers assign.

Starch and Elliott (1912) conducted a study to determine the reliability of the grades assigned by teachers in grading high school English. A set of questions with two answer papers were sent to 200 high schools which were members of the North Central Association in the United States with the request that the principal teacher of first-year English grade the papers according to the practices and standards of the school. It was found that the range of marks given by different teachers to the same paper was as large as 40 percentage points. Furthermore, both papers were not marked correspondingly high or low. Even though paper B was marked on the average eight points lower than paper A, 19 of the teachers marked it higher than paper A, while 23 of the teachers marked it 15 or more points lower than paper A. It was also found that small high schools tended to mark somewhat more leniently than large high schools, but the range of variation was the same for both.

In a later study Starch and Elliott (1913) investigated the reliability of teacher grading in History. Again 200 high schools located in the American Mid-West were sent a set of questions and a copy of an answer paper, with a request that the principal teacher of History grade the paper according to the practices and standards of the school. One hundred and twenty-two of these papers were returned. The distribution of the grades teachers assigned was similar to that found for English with extremes ranging from 43 to 92 percent. Starch and Elliott suggested four major factors that produced such variability of marks:

> (1) Differences among the standards of different schools,
>
> (2) differences among the standards of different teachers,
>
> (3) differences in the relative value placed by different teachers upon various elements in a paper, including content and form, and
>
> (4) differences due to the pure inability to distinguish between closely allied degrees of merit. (p. 681).

The limited body of research, dealing specifically with the criteria upon which teachers base their grades, has focused on the grading of English compositions. Scannell and Marshall (1966) reported a study conducted to determine the relative effects of various error types on the grades assigned by teachers. Five forms of the same answer were developed. The first contained no gross errors, the next contained ten spelling errors, the third contained ten punctuation errors, the fourth contained ten grammatical errors, and the last contained five punctuation, five spelling, and five grammatical errors.

The mean score for the form containing no gross errors was found to be higher than all others and the mean scores were significantly lower at the 0.05 level of significance for the forms containing spelling errors and the form containing a combination of errors.

A similar study was conducted by Marshall (1967) where the primary purpose was to determine the effects of grammatical, punctuation, and spelling errors on teacher grades. The general plan of this study was the same as that used by Scannell and Marshall described earlier. Seven hundred classroom teachers were used as a sample. He found that the means for both the spelling and grammar forms were significantly lower than the means on both the punctuation and combination forms. Marshall also asked teachers four questions concerned with their college background and years of experience in teaching. Marshall reported the results of these questions as follows:

> It was found that when grading essay examinations, beginning teachers are not influenced as greatly as experienced teachers by the presence of a large number of composition errors. Furthermore, it was found that teachers with more training in the content areas tend to grade more severely than do teachers with less training in the area. (Marshall, 1967, p. 386)

Research on Evaluation in Mathematics

As indicated previously, student evaluation in mathematics, and more specifically teacher grading of student work, has received relatively little attention in the literature. However, as Hartung (1949) pointed out, emphasis upon adequate evaluation, both of students and programs is one of the surest and most scientific ways of improving instruction.

The research which has been reported, dealing with evaluation in mathematics, falls into three broad categories: the design and improvement of teacher-made tests, grading practices of teachers related to homework assignments, and the bias of teachers in their grading practices, associated mainly with the sex of the student. A brief description of studies undertaken in these areas follows.

Taylor (1966) studied the effects of varying the instructions of a multiple-choice mathematics test on the achievement of students and the amount of guessing used by students. He found no effect on the mean or variance of the scores for the different types of instruction given to the students. However, the number of items omitted and unfinished suggested that the type of instructions used did have an effect on the amount of guessing.

Hanna (1971) compared three objective item types to determine which would be of most use for testing purposes to replace the formal geometric proof. The first form consisted of statements based on given diagrams and information where the students were asked to determine whether each statement could be proved, disproved, or neither. The second form consisted of multiple-choice questions arranged in pairs. Students were asked to select a multiple-choice response which indicated the 'given' and another which indicated the 'to prove' for a particular situation. The third form consisted of 41 statements divided into four formal proofs where students were asked to select the reason for each statement from a list supplied. A criterion measure was also administered which consisted of four formal proofs to be completed in full. The third form was found to correlate most highly with the criterion test, however, this correlation was not significant.

Coppedge and Hanna (1971) compared the multiple-choice distractors that experienced and student teachers supplied to geometry questions which used numeric and algebraic answers, with the discriminating errors that students actually made in completion format. They found that the single best distractor was identified by teachers only about one-quarter of the time. Furthermore, experienced teachers were slightly less accurate than student teachers.

In the second major area of research various methods of homework grading were compared. Small, Holton, and Davis (1967) compared two methods of homework grading to determine their effect on student achievement. In one group homework was just spot checked once or twice per week, while in the other group papers were checked and graded at least four out of five days per week. The differences between the two groups were not found to be significant but the group which had their papers graded regularly did show slightly greater gains in achievement than the other group.

In a similar study conducted by Austin and Austin (1974), the effect of grading every homework problem with that of grading a randomly selected half of the problems on the achievement of students was compared. Although the overall average achievement was higher for the group having all problems graded, the differences between groups were not significant.

The third area which has received attention in the literature dealing with evaluation in mathematics has been teacher bias in grading practices. Carter (1952) conducted a study to determine whether or not teachers tend to favor one sex over another and whether the sex favored was determined by the sex of the teacher. The general conclusion drawn from this study was that the sex of the teacher was not as important a factor in grade assignment as was the sex of the student. Regardless of whether the teacher was a male or female, boys were penalized in the assignment of grades more than girls, but the penalization was not as great if the teacher was a male.

Fish (1969) studied the relationship of teachers' assigned marks to standardized achievement test results among elementary grade, racially divergent, lower socio-economic status boys and girls. Several subject areas were considered in this study including mathematics. It was found that, despite significant achievement and intelligence differences, teachers in many instances marked high and low achieving pupils in such a way as to fail to distinguish between them. Fish indicated, "the marking practices of teachers ... must be carefully considered and evaluated, so that teachers do not become negative influences on pupil academic behavior." (p. 158)

Duval (1980) conducted a similar study to that of Carter to determine whether teachers applied different grading practices for female students than for male students in mathematics. In the study teachers were asked to grade a geometry final examination paper consisting of four problems presumably written by a high school geometry student. The sex of the student was indicated by a name (Jeanne or Thomas) at the top of the examination paper. Three hundred and fifteen teacher responses were received. The overall finding of this study did not support the contention that secondary teachers were guilty of discrimination between the sexes in the grading of mathematics papers.

The only study found dealing specifically with the reliability of teacher grading in mathematics was carried out by Starch and Elliott (1913). In this study 180 high schools which were members of the North Central Association in the United States were selected. Each school received a set of questions and a copy of the answer paper with the request that the principal teacher of mathematics grade the paper according to the practices and standards of the school. They found an overall range of grades from just under 30 percent to over 90 percent. The variability of grades was found to be even greater for geometry than was the case in the previously discussed studies of Starch and Elliott concerned with the grading of high school history and English. These researchers also examined the amount of variation in the marks assigned to the answer of individual questions. They found that the marks given for a single question varied just about as widely as those of the entire paper.

Summary

As indicated in Chapter I evaluation and measurement are aspects of education whose importance have long been recognized. Students require accurate measurement of their performance to inform them where they stand in relation to others and in relation to mastery of the content. The teacher requires accurate measurement to determine whether or not knowledge is being acquired and to diagnose the strengths and weaknesses of their students. Finally, society requires accurate measurement in order to get a more accurate picture of what each individual has accomplished within the school system.

It was apparent from searching the literature that research in the area of teacher grading in mathematics is limited both in scope and quantity. However, from the work of Starch and Elliott (1913) it was shown that variability in grading does exist, and to a greater extent, in geometry, than high school English, an area where difficulty in achieving grading consistency has long been accepted. Whether the variability of grading in mathematics still exists, and is as pronounced in other branches of mathematics as in geometry, is yet to be determined.

CHAPTER III

METHODOLOGY

In this chapter, the design used in the investigation is described. In the first section the population and sampling procedure are specified. Following this the development, and piloting of the instrument is explained, and the limitations inherent in the design are indicated. Finally, the procedure followed for the distribution of the instrument is explained and the methods used to analyze the data reported.

Population and Sampling Procedure

The population consisted of teachers of the grade 10 Academic Mathematics program in the Province of Newfoundland. To identify the members of this population it was necessary to determine which schools in the Province of Newfoundland offered a grade 10 program of study. Of those schools offering a grade 10 program only one did not offer Academic Mathematics, therefore, this school was not part of the population.

It was determined that 208 schools in the Province of Newfoundland for the school year 1981-82 offered the grade 10 Academic Mathematics program and 100 of these schools were randomly selected as the sample.

Those schools containing students from both the junior and senior high grades and those containing just senior high students with a population of more than 250 were sent three copies of the survey instrument. Similarly, all-grade schools of more than 350 students were sent three copies of the instrument. Schools of smaller size received two copies. These numbers were chosen because it was difficult to determine how many classes of Academic Mathematics were offered in a particular school and whether these classes were all taught by the same teacher or different teachers. Multiple copies of the instrument were distributed to increase the likelihood that more than just senior teachers participated in the survey.

Development of the Test Instrument

The item type selected for grading in this study was the Algebra word problem of a standard textbook type. However, the word problems did differ from the standard textbook wording in that each contained the phrase 'set up and solve an equation'. This phrase was included in order to specify more clearly the objective of the item and thereby eliminate credit for solutions arrived at through other means. The word problem was chosen for two reasons. First of all, it is less algorithmic than many other topics of high school Algebra. Therefore, it is more open to teacher judgments in grade assignment than a purely computational problem or one that just requires algebraic manipulation. Secondly, the word problem is not novel, thus it has a greater likelihood of receiving what can be considered the teacher's typical treatment of a student's answer. Finally, since major changes have recently taken place in the geometry program used in the Province of Newfoundland, making the program itself novel to teachers, any grading practices observed in that area may not be consistent over time.

The initial stage in the development of the instrument was to determine the types of errors students commonly make in solving word problems. Two classes of grade 10 Academic Mathematics totaling 70 students, were requested to supply answers to a set of word problems. Solutions which were totally correct or showed no real attempt were eliminated and the remaining solutions were then grouped, where possible, into categories. Most of the errors could be categorized as, computational, incorrect application of a property, or the use of an incorrect equation. It was found that in a few situations the student managed to obtain a correct numerical solution despite making an error in solving the problem. It was also observed that the same type of error often occurred at different locations in the solution.

Using the information obtained from this pool of student responses a final set of eight categories of student solutions was then developed, using similar student errors so that two problems would exist in each category. The categories selected were as follows:

1) computational error near the beginning of the solution with an incorrect final answer.
2) computational error near the end of the solution with an incorrect final answer.
3) property error near the beginning of the solution with an incorrect final answer.
4) property error near the beginning of the solution with a correct final answer.
5) property error near the end of the solution with an incorrect final answer.
6) property error near the end of the solution with a correct final answer.
7) incorrect equation with incorrect final answer.
8) incorrect equation with correct final answer.

Pilot of Instrument

Each item contained a problem with a solution in which only one error type occurred. Using items from the eight error categories, two forms of the pilot instrument were developed. In form A, teachers were asked to specify in writing why the student lost points and in form B they were asked to check the type of error from a given list.

This phase of the pilot study was carried out for four reasons:

1. To identify any basic weaknesses in the wording or arrangement of the instrument.

2. To determine if the errors made were considered by teachers as realistic and fairly common in occurrence.
3. To determine which of the two formats would be most suitable for use in the main study.
4. To determine if there were other factors that affected the teacher's grade, inherent in the student solutions which were not previously recognized. If such factors were indicated through teacher comments, changes would be made to limit the sources of this variation in grade assignments.

The sample of teachers used for the pilot study was selected from four sources: a city school, a rural school considered large by provincial standards, a smaller rural school, and students enrolled at Memorial University in a high school mathematics methods course who were in their graduating year and had already completed their student teaching requirements. The pilot sample was comprised of four teachers from each of these sources.

Many teachers commented that the errors were realistic and appeared often in their classes. However, one solution was determined to contain an unusual error. Suggestions for improvement were made by two of the participating teachers and the original solution was replaced by one containing the suggested error.

A second change occurred in a problem which requested the student to find the length and width of a rectangle. The student solution originally did now show the student's work in finding the length after the width had been calculated. Since one teacher deducted a point for this omission, the calculation of the length was included in the final instrument.

It was determined that the free response format was most suitable. When given the format of checking the error type, teachers still tended to comment in more detail as to why a particular grade was given. Also, even when the teacher's attention was directed to the selection of the error type, from a list of three, teachers sometimes identified property errors as computational, and vice versa, even though the source of error had been clearly identified by the respondent. Finally, in free response format a more accurate picture of the teacher's reasoning in grade assignments seemed to appear.

Instrument

The final instrument used for this study consisted of a questionnaire containing eight word problems, each accompanied by a sample student response. These responses contained errors which corresponded with the eight previously mentioned categories. The questionnaire also required teachers to specify why the student lost points. This information was requested in order to determine whether teachers were able to identify the error and to also determine if other factors, not previously considered, influenced the grades which a teacher assigned. The instrument is found in Appendix A.

Procedure

The questionnaire for the main study was distributed to the sample group by mail at the end of the third week of April. In a cover letter, to the school principal, the purpose of the study and what the study required of its participants was explained. A copy of this letter is included in Appendix B. A second letter was attached to the rest of the questionnaire supplying more detail to the individual teacher. A copy of this letter is included in Appendix C. A letter was mailed May 12 to act as a reminder to those who had not returned the questionnaire by that time. This letter is included as Appendix D.

A cut-off date for replies was set for June 1. Any responses received after that date were not included in the analysis.

Limitations of the Study

As is the case with all research of this type there are several limitations which were inherent in the design. First of all, teachers may have been defensive about their grading and thus graded the way they thought was expected, rather than their everyday practice. Secondly, teachers are often faced with 30 or 40 papers to grade at one time and may not spend the same time deliberating over one set of answers as would be true in a survey of this type. Thirdly, in the regular classroom situation often other factors such as type of student, student's previous performance, time allotment for testing, or relative difficulty of question, all play a role in the grade a teacher would assign to an answer. In this study the item, isolated from these factors, was to be judged by the teacher. While this eliminated these other variables it may not have been truly reflective of the day to day grading situation. Also, even though two or three copies of the instrument were sent to each school there could have been schools where more than three teachers taught the Academic Mathematics program in grade 10. Therefore, it is possible that just the senior teachers participated in the study in those schools. Furthermore, since more than one copy of the questionnaire was sent to each school, teachers may not have graded independently of each other.

Finally, the inclusion of a question which required the teacher to identify why a particular grade was assigned may have caused the teacher to go back and re-evaluate in light of the question.

Analysis of Data

The item types were arranged for the instrument in the same order as they are presented in the following list. They are labelled I_1 to I_8 for later reference.

- I_1: property error near the beginning with an incorrect final answer.
- I_2: incorrect equation with the correct final answer.
- I_3: computational error near the end of the solution with an incorrect final answer.
- I_4: property error near the end with an incorrect final answer.
- I_5: property error near the beginning with the correct final answer.
- I_6: incorrect equation with the incorrect final answer.
- I_7: computational error near the beginning with an incorrect final answer.
- I_8: property error at the end with the correct final answer.

In this study answers to five basic questions were sought. These questions, as well as the analysis used to test the hypotheses or describe the data, are now discussed.

Question 1: Are teachers as a group consistent in the identification of the errors made by students?

Teachers were asked, after assigning a grade to each word problem, to specify why the student lost points. To answer this question the comments made by teachers were grouped, where possible, under the following categories: reference to specific error type, reference to error type using general terms, incorrect reference or failure to locate error, reference to carelessness, or no comment. When comments were vague other indicators were used to help determine if the error had been successfully located such as the teacher's circling of the line containing the error.

<u>Question 2</u>: Are teachers as a group consistent in the assignment of grades to student responses?

To answer this question the distribution of grades for each item was graphed. In this way a picture of how the grades were distributed for each item was attained. The mean, range and standard deviation for each item were determined and those items with the least and greatest variability identified. Furthermore, individual teacher's grades were examined to illustrate differences in grading patterns that can exist from one individual to another.

In order to answer questions three, four, and five an analysis of variance was carried out based on a repeated measures design. The statistical model used to tabulate and analyse the data is outlined in Table 1.

It is assumed that the scores X_{ij} can be thought of in terms of the following linear model.

$$X_{ij} = \mu + \pi_i + T_j + \pi T_{ij} + E_{ij}$$

where

X_{ij} = an observation on person i under item j.

μ = the grand mean of all potential observations.

π_i = a constant associated with person i.

T_j = the main effect of treatment j.

πT_{ij} = the interaction of person i with treatment j.

E_{ij} = experimental error associated with X_{ij}.

Table 1

Statistical Model

	Treatment (Item Types)									
Person	1	2	3	4	5	6	7	8	Total	Mean
1	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	P_1	$\bar{P}_1$
2	X_{21}	X_{22}	X_{23}	X_{24}	X_{25}	X_{26}	X_{27}	X_{28}	P_2	$\bar{P}_2$
3	.	.	.	.	.	.	.	.	.	.
4	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
i	X_{i1}	X_{i2}	X_{i3}	X_{i4}	X_{i5}	X_{i6}	X_{i7}	X_{i8}	P_i	$\bar{P}_i$
Total	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	G	$\bar{G}$
Mean	$\bar{T}_1$	$\bar{T}_2$	$\bar{T}_3$	$\bar{T}_4$	$\bar{T}_5$	$\bar{T}_6$	$\bar{T}_7$	$\bar{T}_8$		

Statements of questions three, four and five along with the associated hypotheses and the comparisons used to test these hypotheses follow. If the analysis of variance showed no significant differences between scores on items, then answers to questions three, four, and five would not have been pursued further. A nonsignificant result at that point would indicate no significant differences between the grades assigned to any of the eight items.

Question 3. Does the type of error make a difference to teachers' grading of student responses?

To answer this question the following hypothesis was tested:

> Hypothesis 1: There is no significant difference between the grades that teachers assign to property errors and computational errors.

To test this hypothesis three comparisons were considered. The first comparison involved two items, one having a property error and the other a computational error, where both errors were located near the beginning of the solution, resulting in an incorrect numerical answer. The second comparison involved two items, one having a property error and the other a computational error where both errors were located near the end of the solution, resulting in an incorrect numerical answer. In the final comparison, the items containing the property error were combined and this combination compared with the combined computational errors. These comparisons were as follows:

(a) $H_o: \mu_7 - \mu_1 = 0$

(b) $H_o: \mu_3 - \mu_4 = 0$

(c) $H_o: (\mu_7 + \mu_3) - (\mu_1 + \mu_4) = 0$

Question 4. Does the location of an error make a difference in the teachers' grading of student responses?

To answer this question the following hypothesis was tested:

> Hypothesis 2: There is no significant difference between the grades teachers assign when an error is made at the beginning or end of a solution.

To test this hypothesis four comparisons were considered. The first comparison involved two items having computational errors, one near the beginning and the other near the end of the solution, where both resulted in an incorrect numerical answer. The second comparison involved two items having property errors, one near the beginning and the other near the end of the solution, where both resulted in an incorrect numerical answer. The third comparison involved two items having property errors, one near the beginning and the other near the end of the solution, where both resulted in a correct numerical solution. The final comparison involved the combination of the three previously mentioned pairs of items. These comparisons were as follows:

(a) H_o: $\mu_3 = \mu_7 = 0$

(b) H_o: $\mu_4 - \mu_1 = 0$

(c) H_o: $\mu_8 - \mu_5 = 0$

(d) H_o: $(\mu_3 + \mu_4 + \mu_8) - (\mu_7 + \mu_1 + \mu_5) = 0$

Question 5: Does the correct answer make a difference to teachers' grading of student responses?

To answer this question the following hypothesis was tested:

Hypothesis 3: There is no significant difference between the mean grades given by teachers when the answer is incorrect or correct.

To test this hypothesis four comparisons were considered. The first comparison involved two items, both having an incorrect equation where one resulted in a correct and the other an incorrect numerical answer. The second comparison involved two items both having property errors near the end, where one resulted in a correct and the other an incorrect numerical answer. The third comparison involved two items both having property errors near the beginning, where one resulted in a correct and the other an incorrect numerical answer. The final comparison involved the combination of the three previously mentioned pairs of items. These comparisons were as follows:

(a) H_o: $\mu_2 - \mu_6 = 0$

(b) H_o: $\mu_8 - \mu_4 = 0$

(c) H_o: $\mu_5 - \mu_1 = 0$

(d) H_o: $(\mu_2 + \mu_8 + \mu_5) - (\mu_6 + \mu_4 + \mu_1) = 0$

All comparisons were tested at the 0.01 level of significance using the Scheffé method of multiple comparisons.

CHAPTER IV

ANALYSIS OF DATA

The main purpose of this study was to investigate the grading practices of mathematics teachers. The content area selected was word problems from grade 10 Algebra. In this chapter the data, which were collected according to the procedures outlined in Chapter III are examined in terms of the stated questions.

The population considered for the study consisted of all teachers of grade 10 Academic Mathematics in the province of Newfoundland for the school year 1981-82. Of the 100 schools selected, 69 responded prior to the deadline of June 1, resulting in a total sample of 90 teachers.

Question 1

Are teachers as a group consistent in the identification of errors made by students?

To answer this question the comments made by teachers on each item were grouped, where possible, under the following categories: reference to specific error type, reference to error type using very general terms, incorrect reference to error or failure to locate error, reference to sloppy or careless work, or no comment. If comments were omitted or difficult to interpret, other indicators such as circling the line containing the error, or use of an arrow to indicate error location were used. This information formed the basis for determining the consistency of the teachers in the identification of errors.

It was found that on the whole teachers were very consistent in their identification of student errors. There was, however, some variation in the terms used by these teachers to classify the errors, but only in a few isolated cases were the errors not located or located incorrectly.

Item 1 contained a property error near the beginning of the solution resulting in an incorrect final answer. The exact property error was the improper use of the distributive property. Fifty-nine of the teachers surveyed specifically identified the error as the incorrect use of the distributive property. Twenty referred to the error in more general terms such as computational, arithmetic, or lacking the ability to multiply. Eight referred to the error as reflecting careless or sloppy work on the part of the student. Three teachers supplied no comment. Even when the comments differed greatly it was clear that teachers were able to locate the source of error.

Item 2 involved the use of an incorrect equation where the student obtained a correct numerical solution. Eighty-six of the teachers identified the error to be concerned with the incorrect equation. This group most commonly referred to the error as a lack of understanding of the distance, rate, and time relationship or by simply referring in general terms to the incorrect equation. Two of the respondents did not specify the reason for the students loss of points and two did not identify the equation as incorrect, but instead, determined the error to be located at some other point in the solution. One respondent felt the solution lacked detail and the other indicated that the error was located in line two but did not specify the nature of the error.

Item 3 contained a computational error near the end of the solution. This computational error involved the incorrect multiplication of an improper fraction by a whole number. Seventy-seven of the teachers described the error using such terms as multiplication error, mechanical error, computational error or lack of basic skills. Ten referred to the error as careless. Two referred to the error as algebraic and two supplied no comment, however, indicators such as circling or using an arrow to indicate the error, were used to show that the line containing the error had been located. One of the teachers felt the equation was incorrect, indicating that the equation should have read $2/3\ (x + 11)$ instead of $2/3 \cdot x + 11$.

Item 4 contained a property error located near the end of the solution, resulting in an incorrect answer. The exact property error was the adding of unlike terms. Fifty-three of the teachers referred to the error specifically as relating to an incorrect combination of terms. Twenty-six referred to it in more general terms such as property error, or algebraic error, and four described the error as reflecting sloppy or careless work. Five of the teachers termed the error as computational and two supplied no comment. All seemed to have correctly located the source of error.

Item 5 contained a property error located near the beginning of the solution where the student obtained a correct numerical answer. The exact property error was the incorrect application of the order of operations principle. Seventy-one of the teachers identified the error specifically as relating to the use of the order of operations, while eight referred to the error using more general terms such as a property or algebraic error, and two termed the error carelessness. Six teachers indicated the error as relating to a lack of understanding of the distributive property, while two supplied no comment. One teacher determined that the equation was incorrect but did not indicate what the correct equation should be.

Item 6 involved the use of an incorrect equation due to improper use of the relationship between area, length, and width of a rectangle. Eighty-six of the teachers identified the error to be concerned with the incorrect equation. Four of the teachers did not specify why the grade was assigned. It appeared that all teachers had successfully located the error.

Item 7 contained a computational error located near the beginning of the solution. The specific error involved the incorrect multiplication of a rational number by a whole number. Eighty-four of the teachers referred to the error using one of the following terms, multiplication error, computational error, error in arithmetic or mechanical error. Two determined the error to be related to the incorrect use of the distributive property and one referred to the error as dealing with a mistake in subtraction. Three teachers supplied no comment. All appeared to have successfully located the error.

Item 8 contained a property error located near the end of the solution which resulted in a correct numerical answer. The exact property error was the combining of unlike terms by addition. Fifty-three of the teachers referred specifically to the incorrect combination of terms and 33 used more general terms such as mechanical error, property error, or algebraic error, while two termed the error careless. One found no error in the solution and one made no comment.

Except for a few isolated instances teachers were very consistent on all items in the location and identification of errors. While the terminology used to refer to the errors varied greatly, in the vast majority of cases it was evident that the error had been successfully determined.

Question 2

Are teachers as a group consistent in their assignment of grades to student responses?

The answer to this question was sought in order to determine any variation that existed in teacher grade assignments. Since the combination of item scores might result in a high grade assignment on one item balancing out a low grade assignment on another item, the distribution of grades for each item was considered. Then, other factors such as severity, amount of grade distribution throughout the grading scale and relative judgments made by teachers were considered.

In Figure 1 the distribution of grades assigned by teachers on item 1 is presented. The student error involved the incorrect use of the distributive property near the beginning of the solution. The grades assigned to this item ranged from two to 10, out of 10. The mean score for the item was 6.28 with a standard deviation of 1.63.

In Figure 2 the distribution of grades assigned on item 2 is presented. The student error involved the incorrect equation due to improper use of the distance, rate, and time relationship, with the correct numerical solution being obtained. Some teachers expressed strongly the opinion that since the equation was incorrect the student should receive no credit for any part of the solution. There were also several who indicated that the correct identification of the variables and representation

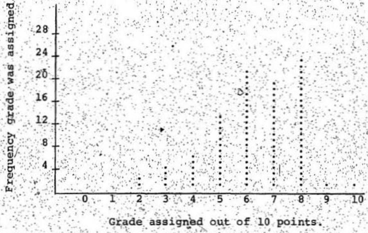


Figure 1. Distribution of grades assigned to item 1.

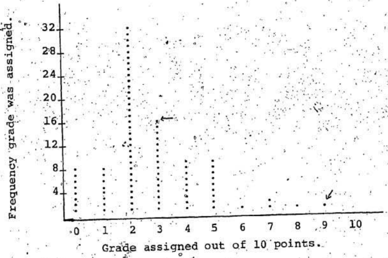


Figure 2. Distribution of grades assigned to item 2.

Note: The arrows indicate a grade assigned by an individual who did not identify the error, or identified it incorrectly.

of the problems information in diagram form reflected some understanding and deserved partial credit. The grades assigned to this item ranged from zero to nine out of 10, but since the nine out of 10 was assigned by an individual who had not identified the error correctly, the true range for this error type can be considered zero to eight out of 10. The mean score of the item was 2.77 with a standard deviation of 1.82. The variability of grading on this item was the second highest of all eight items.

In Figure 3 the distribution of grades on item 3 is presented. The student error was computational, specifically the incorrect multiplication of a fraction by a whole number, the error being located near the end of the solution.

Some teachers expressed the view that this error reflected carelessness on the part of the student rather than lack of understanding, therefore, they felt the student should not be penalized severely for the error. Grades on this item ranged from one to 10 out of 10, but since the one out of 10 was assigned by an individual who had not identified the error correctly, the true range for the error type can be considered six to 10. The mean score for the item was 8.41 with a standard deviation of 1.07. This item had the least variability and the highest mean of all eight items.

In Figure 4 the distribution of grades assigned to item 4 is presented. The error involved the adding of unlike terms, located near the end of the solution. Grades

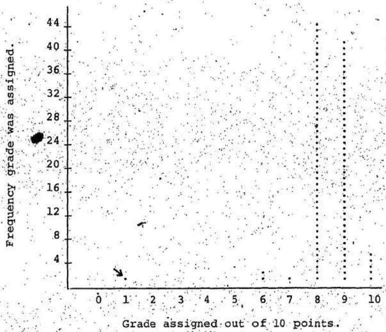


Figure 3. Distribution of grades assigned to item 3.

Note: The arrows indicate a grade assigned by an individual who did not identify the error, or identified it incorrectly.

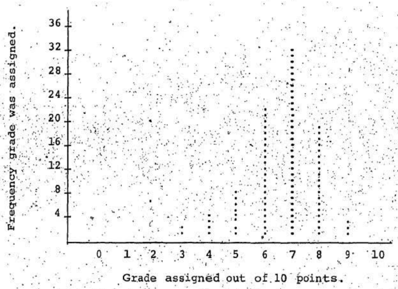


Figure 4. Distribution of grades assigned to item 4.

assigned to item 4 varied from three to nine out of 10. The mean score on the item was 6.63 with a standard deviation of 1.26. The grading of this item reflected much greater consistency than most of the items, having just slightly more variability than the computational error made at the end of the solution.

In Figure 5 the distribution of grades on item 5 is presented. An error was made in applying the order of operations principle at the beginning of the student solution. This resulted in a correct numerical solution.

Grades on this item varied from zero to 10 out of 10. It should be noted that the zero was assigned by a teacher who identified the error differently than the rest of the subjects. This teacher determined the equation itself to be incorrect. The mean score on this item was 5.40 with a standard deviation of 1.97. The grading of this item reflected the greatest variability of all items. Some teachers commented that they considered the order of operations a very important basic principle and an error in its use should be considered serious. A few teachers also commented that the correct numerical solution had no affect on the grade they assigned.

In Figure 6 the distribution of grades assigned to item 6 is presented. The student error involved the use of an incorrect equation due to improper use of the formula used to find the area of a rectangle. Comments on this

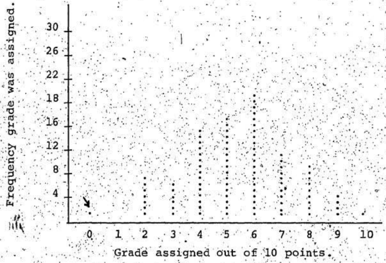


Figure 5. Distribution of grades assigned to item 5.

Note: The arrows indicate a grade assigned by an individual who did not identify the error, or identified it incorrectly.

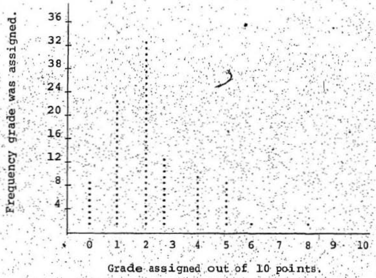


Figure 6. Distribution of grades assigned to item 6.

item were similar to that on item 2 where an incorrect equation was also formed.

Grades assigned to this item ranged from zero to eight out of 10. The mean score for the item was 2.23 with a standard deviation of 1.51. This item had the lowest mean score of all eight items.

In Figure 7 the distribution of grades assigned to item 7 is presented. The student error was computational, more specifically an error in the multiplication of a rational number and a whole number, located near the beginning of the solution. Grades assigned to this item ranged from three to 10 out of 10. The mean score for the item was 7.28 with a standard deviation of 1.69. The computational error located near the beginning produced a much greater variability in grades than did a similar error occurring near the end of the solution.

In Figure 8 the distribution of grades assigned on item 8 is presented. The student error was made in the adding of unlike terms. The error was located near the end of the solution, but resulted in a correct numerical answer.

Grades assigned for this item ranged from three to 10 out of 10. The mean score was 6.86 with a standard deviation of 1.42. Teachers generally commented that the correct answer had no effect on the grade they assigned to the item. Variability of grading on the item was low in comparison to most of the other items.

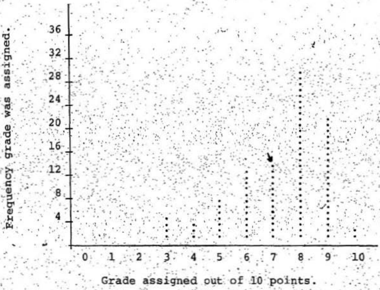


Figure 7. Distribution of grades assigned to item 7.

Note: The arrow indicates a grade assigned by an individual who did not identify the error, or identified it incorrectly.

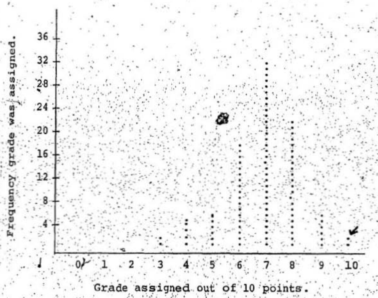


Figure 8. Distribution of grades assigned to item 8.

Note: The arrow indicates a grade assigned by an individual who did not identify the error.

In Table 2 the descriptive statistics related to the eight items are presented. Even though this information has been previously mentioned, a summary of the relevant statistical information relating to question 2 is presented here in tabular form for easy reference.

Table 2

Range, Mean, Standard Deviation of Grades on Each Item

Item	Number	Max Value	Min Value	Mean	Standard Deviation
1	90	10	2	6.28	1.63
2	90	9	0	2.77	1.82
3	90	10	1	8.41	1.07
4	90	10	3	6.63	1.26
5	90	10	0	5.40	1.97
6	90	8	0	2.23	1.51
7	90	10	3	7.28	1.69
8	90	10	3	6.86	1.42

To illustrate further the types of inconsistencies that occur in grading, individual subjects were selected. By graphing the grades assigned by these subjects, other types of grading inconsistency were exemplified.

Thorndike (1971) indicated three ways that grading practices may vary. First of all, graders differ in severity. In examining differences in severity of teacher grades the total grade assigned for all eight items for each teacher was considered. These grades ranged from 24 to 62 out of 80, having a mean of 45.9 and a standard deviation of 7.58. To demonstrate differences in severity, three subjects who appeared to exhibit similarities in ranking the importance of errors and differences in the total assigned grade were selected. The grades given by these subjects on each item are shown in Figure 9. It is evident that subject A was more severe than subject B and subject B was more severe than subject C.

Thorndike also pointed out that teachers differ in the extent to which they distribute grades throughout the rating scale. Some teachers tend to group grades closely around the mean while others spread them throughout the scale.

In Figure 10 the grades given by two subjects illustrating differences in grade distribution throughout the scale are diagrammed. The mean grade assigned by subject D was 6.50, and the mean grade assigned by subject E was 6.65. Even though the mean scores were very similar the standard deviation of the two sets of scores differed greatly. The grades of subject D had a standard deviation of 3.46, while the grades of subject E had a standard deviation of 1.41.

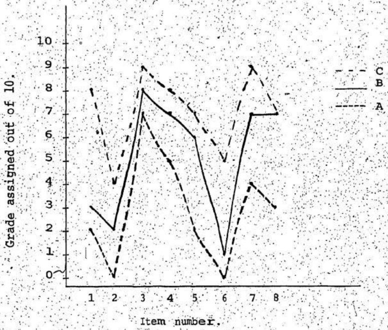


Figure 9. Comparison of subjects to illustrate differences in severity.

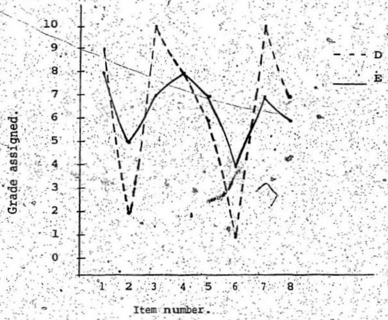


Figure 10. Comparison of subjects to illustrate differences in amount grades were distributed throughout the grading scale.

The third difference in grading that Thorndike discussed pertained to the differences in relative value teachers assign to different items. In Figure 11 the grades assigned by subjects F and G are illustrated. It can be discerned from the graphs of the grades assigned, that the two individuals differ in their relative judgment of importance of the errors contained in the eight items.

Analysis of Variance

Before answering questions three, four and five an analysis of variance was carried out based on a repeated measures design. The results of the analysis of variance are shown in Table 3.

Table 3

Results of Analyses of Variance

Source	SS	DF	MS	F
Between people	639.89	89	7.19	
Within people	4095.37	630	427.15	
Item	2977.54	7	425.36	237.63*
Residual	1117.83	623	1.79	

* F-ratio significant at the 0.01 level.

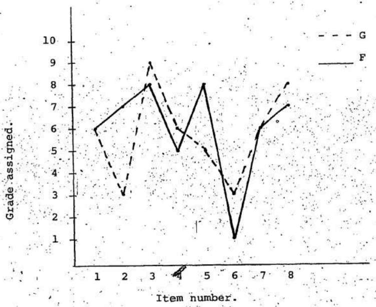


Figure 11. Comparison of subjects to illustrate difference in judgments of relative importance of error.

Having established that the overall analysis of variance was significant the comparisons associated with questions 3, 4 and 5 were examined. The critical value at the 0.01 level of significance for all comparisons was 6.63.

Question 3

Does the type of error make a difference to teachers grading of student responses?

In answering this question the following hypothesis was tested.

Hypothesis 1. There is no significant difference between the grades teachers assign to property errors and computational errors.

In testing hypothesis 1, three comparisons were made. The results of these comparisons are reported in Table 4.

Table 4

Comparisons Associated With Hypothesis 1

Number	Comparison	Mean Difference	F-value
1	$\bar{X}_{.7} - \bar{X}_{.1}$	1.00	25.08*
2	$\bar{X}_{.3} - \bar{X}_{.4}$	1.78	79.45*
3	$(\bar{X}_{.7} + \bar{X}_{.3}) - (\bar{X}_{.1} + \bar{X}_{.4})$	2.78	96.90*

* F-ratio significant at the 0.01 level.

In the first comparison, reported in Table 4, a property error was compared with a computational error, both made near the end of the solution. The second comparison involved a property error and a computational error, both made near the beginning of the solution. The third comparison combined the items containing the property errors and compared this combination with the combined computational errors. In all three cases the differences were found to be significant at the 0.01 level with the property errors being graded more severely than the computational errors.

Question 4

Does the location of an error make a difference in the teachers grading of student responses?

To answer this question the following hypothesis was tested:

> Hypothesis 2: There is no significant difference between grades teachers assign when an error is made at the beginning or end of a solution.

In testing hypothesis 2, four comparisons were made. The results of these comparisons are reported in Table 5.

In the first comparison reported in Table 5 a computational error made near the beginning is compared with a computational error made near the end of the solution, both resulting in an incorrect numerical answer. The second comparison was made between a property error made near the beginning and a property error made near the end of the solution, both resulting in an incorrect numerical answer. The third comparison was made between a property error near the beginning and a property error near the end of the solution, both resulting in a correct numerical answer. Finally, comparison four involved the combination of the three previous items having errors made near the beginning with the three items having errors made at the end. Three of the four comparisons were significant. The error made at the beginning was graded more severely than the error made at the end. Comparison two, although not significant, did show a lower grade for the error made at the beginning.

Table 5

Comparisons Associated With Hypothesis 2

Number	Comparison	Mean Difference	F-value
1	$\bar{X}_{.3} - \bar{X}_{.7}$	1.13	32.39*
2	$\bar{X}_{.4} - \bar{X}_{.1}$	0.35	3.18
3	$\bar{X}_{.8} - \bar{X}_{.5}$	1.46	53.26*
4	$(\bar{X}_{.3} + \bar{X}_{.4} + \bar{X}_{.8}) - (\bar{X}_{.7} + \bar{X}_{.1} + \bar{X}_{.5})$	2.94	72.65*

* F-ratio significant at the .01 level.

Question 5

Does the correct answer make a difference to teachers grading of student responses?

To answer this question the following hypothesis was tested:

Hypothesis 3: There is no significant difference between the mean grades given by teachers when the answer is correct or incorrect.

In testing hypothesis 3, four comparisons were made. The results of these comparisons are reported in Table 6.

Table 6

Comparisons Associated With Hypothesis 3

Number	Comparison	Mean Difference	F-ratio
1	$\bar{X}_{.2} - \bar{X}_{.6}$	0.54	7.15*
2	$\bar{X}_{.8} - \bar{X}_{.4}$	0.23	1.24
3	$\bar{X}_{.5} - \bar{X}_{.1}$	-0.88	19.37*
4	$(\bar{X}_{.2} + \bar{X}_{.8} + \bar{X}_{.5}) - (\bar{X}_{.6} + \bar{X}_{.4} + \bar{X}_{.1})$	-0.11	0.13

* F-ratio significant at the .01 level.

In the first comparison reported in Table 6 two items, both having an incorrect equation, where one resulted in a correct and the other an incorrect numerical answer were considered. This comparison resulted in a significant difference, where the item containing the incorrect numerical answer was graded more severely than the item with a correct numerical answer.

The second comparison was made between two problems, both having a property error near the end, one resulting in a correct and the other an incorrect numerical answer. The difference in grading on these items was not found to be significant, however, grades assigned to the item having a correct numerical answer were slightly higher than those assigned to the item having an incorrect numerical answer.

The third comparison was made between two problems, both having a property error near the beginning, where again one resulted in a correct and the other an incorrect numerical answer. This comparison resulted in a significant difference where the item containing a correct answer was graded more severely than the item with an incorrect numerical answer.

Finally, comparison four combined the three previous items which resulted in a correct numerical answer with the three resulting in an incorrect numerical answer. The difference was not found to be significant for this overall comparison.

CHAPTER V

DISCUSSION, IMPLICATIONS AND RECOMMENDATIONS

Overview

In this study the grading practices of teachers using grade 10 Algebra word problems were examined. In examining these grading practices the consistency of teachers in the identification, classification, and grading of student answers was investigated. Furthermore, an attempt was made to identify the effects of error type, error location, and whether the student obtained a correct or incorrect numerical answer, on the grading practices of teachers.

The teachers chosen to participate in the study were randomly selected from the population of all teachers of grade 10 Academic Mathematics in the Province of Newfoundland for the school year 1981-82. One hundred schools, offering the Academic Program at the grade 10 level, were selected to participate.

The research technique employed to conduct the study involved the use of a survey questionnaire. This questionnaire consisted of eight word problems, each followed by a student solution. Teachers were asked to grade each item out of 10 points and specify why the grade was assigned.

Discussion of Results

The results of this study were presented in Chapter IV. In this section these results are discussed.

The first question which this study sought to answer concerned teacher consistency in identification of errors made by students. The purpose of this question was to ascertain whether all teachers were identifying, and thus grading, the same error. If many teachers failed to successfully locate an error, the grades they assigned to that item would not be comparable.

As presented in Chapter IV, the number of teachers who failed to locate the errors were very few in number and no individual located more than one incorrectly. In fact, of the 720 items graded, there were only five instances of teachers missing the error or determining it to be different from the actual error. It was assumed that in instances where the error was not successfully located the grader was making a purely accidental oversight not reflecting a lack of ability, since errors were successfully located 99.3% of the time.

In identifying the errors, teachers used both general and specific classification terminology. For example, item one contained an error in the use of the distributive property and this was referred to by some teachers specifically as a distributive error while others referred to the error using such general terms as algebraic or property error. In other cases an error type was referred to using an inappropriate term. For example, in item 7 a multiplication error was referred to by one individual as an error in subtraction, however, no one individual made this type of mistake with more than one item. Most often teachers described the error using either exact terminology or more global categorizations into which the specific error type fell.

It is deserving of note the number of times teachers referred to student errors as careless. While this comment occurred most often with an item containing a computational error, it also appeared fairly frequently when the error involved the incorrect use of a property. However, carelessness was not mentioned in reference to either of the two items which had an incorrect equation formed. Even though careless errors are made by students, this reason is perhaps too easily relied upon by teachers and can result in a neglect of the student's real problem.

The second question the study attempted to answer was related to the consistency of teachers in the assignment of grades to student answers. To answer this question teachers were asked to assign a numerical grade out of 10 points to each item. The data were then analyzed from several different perspectives.

First of all, the grades assigned to each item were graphed to determine variability for each error type.

Generally, grades were found to be least variable when errors occurred near the end of the solution. An item with a computational error near the end had the lowest variance followed by two items containing property errors near the end of the solution. Grades were found to be more variable when errors were located near the beginning of the solution.

It appeared through close inspection of the grading techniques that the approach to grading fell into two major categories. The first category involved those who had set up a marking scheme, outlining the relative weight given to each component of the student's solution. A common system included awarding one to two points for the correct identification of variables, three to six for setting up the correct equation, two to four for solving the equation and sometimes one point for writing a sentence to report the final answer.

The second category involved those who graded in a linear sequence where points were assigned according to how far the student reached in the solution process before an error was made. Some teachers actually pointed out that this was their approach, while others used methods such as correcting each line of the solution, or placing one point next to each correct line. These were used as indicators of linear grading. One teacher commented that it was difficult to grade a solution out of ten points when the solution only consisted of nine lines.

These very different approaches to the grading process may account for some of the differences in variability that occurred between errors made at the beginning and end of the solution. For example, a grader from either group would probably judge a computational error at the end as a fairly trivial error and thus assign a reasonably high grade. However, the person who grades in a linear sequence may well assign a lower grade when the error is made early in the solution. This would cause a greater spread for errors at the beginning than errors at the end.

Thorndike (1971) identified three reasons why consistency of grading was difficult to achieve in essay grading. To determine if these reasons were applicable to the word problems used in this study the data were analyzed to give further insight into Thorndike's statements. First of all, Thorndike indicated raters differed in severity. One means of determining severity was to consider the total grades assigned by individual teachers. It was found that the totals varied from 30% to 77.5%.

To illustrate further differences in severity, the grades assigned by individual subjects who showed similarity in ranking the importance of errors were examined. From examining the grades assigned by individual subjects it was apparent that teachers differed in the level of severity which they maintained. That is, some teachers consistently assigned lower grades than others. Several possible explanations may account for such differences. Some teachers are naturally more lenient than others. Ebel (1972) indicated that instructors are often under pressure to give more high and fewer low marks. This pressure can be the result of a wish to avoid controversy, the establishment of a favorable view of the course or of the teacher by students, or fear that the administration will equate lower marks with poor teaching.

The second reason Thorndike identified related to the extent to which teachers distribute their grades throughout the grading scale. To demonstrate the pertinence of this reason to the results of this study, grades assigned by individual subjects were again examined. From examining the grades assigned by individual subjects it was apparent that differences did exist in the way teachers distributed their grades throughout the grading scale. Many of the pressures to give higher marks outlined earlier may also be responsible for lower grade distributions. A proliferation of high grades is looked at just as suspiciously as too many low grades, therefore, the individual who feels insecure, or is new to the school or to teaching, may seek to avoid controversy by grouping grades close to the mean, thus avoiding the assignment of very high or very low grades.

The final reason identified by Thorndike was concerned with the relative value teachers place on student work. This was illustrated to a certain extent in the discussion of the distribution of grade assignments to particular items. To emphasize teacher differences in relative values assigned to item types, grades assigned by individual subjects were again examined. From examining the grades assigned by individual subjects it was also apparent that teachers differ in the relative values they assign to various error types. Thorndike indicated that such diversity of judgment can be explained to some extent as, "reflecting only random differences that accompany any human response: if so, they are inescapable." (p. 277) Other possible explanations for individual differences in relative judgments may relate to background of the grader in terms of experience and education.

Evidence of grading inconsistency was identified on several different levels. Teachers varied greatly in the grades they assigned to various error types, they differed in severity, grade distribution throughout the grading scale and in relative judgment of importance of various error types.

In question 3 an attempt was made to determine whether or not the type of error made a difference to teacher grading of student answers. Only two error types were considered, the property error and the computational error. Significant differences were found, with teachers grading property errors more severely than computational errors. Property errors are probably graded more severely because they are more easily interpreted to reflect conceptual difficulty. Especially with calculator use, computational errors are more difficult to interpret in relation to student understanding or ability. It may, therefore, be reasonable to postulate that when points are subtracted for computational errors the teachers aim is to either point out carelessness, stress student checking of work, or emphasize mathematical exactness.

In question 4 the effects of error location on grading were considered by comparing an error made at the beginning with an error made at the end of a solution. Three of the four comparisons showed significant differences, where teachers graded errors made at the beginning more severely than errors made at the end. This can, perhaps, again be attributed to the two major approaches to grading that were discussed earlier. Teachers who grade in a linear sequence, would assign higher grades to solutions having the error located near the end. In a few instances, even when teachers set up a marking scheme, the actual solution of the equation again appeared to be graded in a linear fashion, once more causing lower grades for errors made at the beginning.

In question 5 an attempt was made to determine whether or not the correct or incorrect final answer would make a difference to teacher grading, when an error had occurred in the solution. The answer to this question was not conclusive. Of the three pairwise comparisons, two were in the direction favoring the correct numerical answer with one of the two showing significant differences. Another comparison was also significant but this time in the direction favoring the incorrect final answer. This may have been caused by an interaction effect with the error types used in the two items considered. Of the two items compared, one involved the incorrect use of the distributive property and the other the incorrect use of the order of operations principle, both located near the beginning of the solution. If teachers judged one of these error types to be of greater importance, the severity of grading on that item would have been affected.

For the overall comparison, in which all items having the correct numerical solution were compared with those having an incorrect numerical solution, the difference was not significant. In grading many papers or in rushed circumstances the correct answer might result in some teachers assigning a higher grade but no overall relationship was discernible through this study.

Implications of Results

The results of this study have several implications for educators. The general conclusion was that inconsistencies do exist in teacher grading practices on several different levels, including judgments of relative importance of error types, level of severity, and distribution of grades throughout the grading scale. As pointed out in Chapter I, such inconsistencies can result in a student failing in one school, with a particular teacher while the same student could conceivably see a measure of success in another school or with another teacher. When the grades that teachers assign are used for such a variety of purposes on the instructional and administrative level, it is important that teachers realize the extent of existing grading inconsistencies.

The significant differences found in the location of errors is a particularly troublesome finding. That location should make such differences, both in the mean grade and the variance of grades, indicates that perhaps too many teachers are grading by the use of linear sequence grading methods, a practice that should be considered questionable in light of the difference in scores resulting from this practice.

While there is no set, carved in stone, means of grading, it is important that the grader consider carefully what he or she wishes the student to accomplish, what objective each item is testing and whether the student has reached that objective. When point breakdown is necessary or desirable, marking schemes should be established, decisions on relative weight of the components should be made and the importance of exactness for the particular item type decided upon. As indicated by Thorndike and Hagan (1977),

> Like any deeply ingrained aspect of a culture, grading, and marking practices involve motivations that are only partly accessible to inspection at the conscious level, and are correspondly resistant to change. Before practices can be made psychometrically more sound, their biases must be brought out and subjected to conscious, even self-conscious scrutiny.

> Since many, perhaps most, of the issues relating to the assignment of marks are issues of value, an examination of marking practices must address itself first and foremost to a clarification of these values. (p. 607)

While in this study just a few of the many factors which can affect the grading practices of teachers were examined the results should aid in the illustration of the inconsistencies which exist and some of the sources of these inconsistencies. As Ebel (1972) indicated, the beginning of wisdom in marking is the recognition of its shortcomings and the cultivation of that wisdom is to work at improving them.

Recommendations for Further Study

This study has, in many respects, been a pilot study since so little work has been done in the area. Its purpose was to determine some of the inconsistencies that presently exist, but not to offer means of reducing these inconsistencies. Just a few of the factors which influence teacher grading practices were examined and those over a rather narrow content area and population. However, as a result of the study several recommendations for research can be made.

A similar study could be conducted which focuses on the same factors, using teachers of other grade levels or mathematics courses to determine if factors such as error location and type effect these teachers' grades differently. Also, different data collection techniques could be used, such as personal interviews, so that the exact 'formula' applied by teachers in grading could be more clearly ascertained.

It would also be worthwhile to examine more closely the grading patterns applied by individual teachers. In this study two basic patterns of grading seemed to emerge. Perhaps the approaches to grading applied by teachers could be sub-divided and examined in more detail.

The three reasons for inconsistencies in grading identified by Thorndike and discussed in some detail in this report may warrant further research to determine the extent to which these reasons apply to grading in mathematics.

A study to determine the relationship of certain teacher characteristics to the grades these teachers assign might also provide useful information. Such characteristics might include sex of the teacher, number of years teaching experience, and background in mathematics and mathematics education.

A study to determine common grading schemes presently in use by teachers with the aim of developing specific grading schemes to provide guidelines, especially for new teachers, in grading particular types of content, would also be worthwhile.

Furthermore, a limitation of this study was concerned with the isolation of the item being graded from the many factors which occur in the regular classroom situation that often effect grading. Therefore, a study conducted within the classroom situation, to determine grading practices may also prove worthwhile.

BIBLIOGRAPHY

Bibliography

Adams, Georgia Sachs and Torgerson, Theordore L. Measurement and Evaluation in Education, Psychology and Guidance, Holt, Rinehart and Winston, New York, 1964.

Ahmann, J. Stanley and Glock, Marvin D. Evaluating Student Progress. Allyn and Bacon, Inc., Boston, 1981.

Austin, Joe Dan and Austin, Kathleen A. Homework grading procedures in junior high mathematics classes, School Science and Mathematics, 1974, 74(4), pp. 269-272.

Carter, Robert Scriven. How invalid are marks assigned by teachers? Journal of Educational Psychology, 1952, 43, pp. 218-228.

Coppedge, Floyd and Hanna, Gerald S. Comparison of teacher-written and empirically derived distractors to multiple-choice test questions. School Science and Mathematics, 1971, 71(9), pp. 299-303.

Course Description, Academic Mathematics 1203, Division of Instruction, Department of Education, Government of Newfoundland and Labrador, 1981.

Davis, Fredrick B. Educational Measurements and Their Interpretations. Wadsworth Publishing Company, Inc., Belmont, California, 1964.

Durost, Walter N. and Prescott, George A. Essentials of Measurement for Teachers. Harcourt, Brace, and World, Inc., New York, 1962.

Duval, Concetta M. Differential grading behavior towards female students of mathematics. Journal for Research in Mathematics Education, 1980, 11(3), pp. 202-214.

Ebel, Robert L. Measuring Educational Achievement, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1965.

Ebel, Robert L. Essentials of Educational Measurement, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1972.

Fish, Enrica. The relationship of teachers' assigned marks to tested achievement among elementary grade, racially divergent, lower socio-economic status boys and girls. University of Minnesota, Ph.D., 1969.

Gaier, Eugene L. The Grade Society. Knapp, Thomas R. and Miller, John K. Readings in Statistics, Measurement, and Research, Simon and Schuster, New York, 1970.

Gronlund, Norman E. Measurement and Evaluation in Teaching. The Macmillan Company, New York, 1971.

Hadley, Trevor S. A School Mark - Fact or Fancy? Educational Administration and Supervision, 1954, 40, pp. 305-312.

Hanna, Gerald S. Testing students' ability to do geometric proofs: a comparison of three objective item types. Journal for Research in Mathematics and Education, 1971, 2(3), pp. 213-217.

Hartung, Maurice L. A forward look at evaluation. The Mathematics Teacher, 1949, 62(1), pp. 29-33.

Hill, John R. Measurement and Evaluation in the Classroom. Charles E. Merrill Publishing Company, Columbus, Ohio, 1976.

Kulm, Gerald. The effects of two summative evaluation methods on achievement and attitudes in individualized seventh-grade mathematics, School Science and Mathematics, 1977, 77(8), pp. 639-647.

Lien, Arnold J. Measurement and Evaluation of Learning. William C. Brown Co. Publishers, Dubuque, Iowa, 1976.

Majors, Gene W. and Michael, Joan J. The relationship of achievement on a teacher-made mathematics test of computational skills to two ways of recording answers and to two workspace arrangements. Educational and Psychological Measurement, 1975, 35(2), pp. 1005-1009.

Marchall, J.C. Composition errors and essay examination grades reecamined. American Educational Research Journal, 1967, 4, pp. 375-386.

Pfeiffer, Isobel and Davis, O.L. Teacher-made examinations: What kind of thinking do they demand? Bulletin of NASSP, 1965, 49, p. 302.

Scannell, Dale P. and Marchall, Jon C. The effects of selected composition errors on grades assigned to essay examinations. American Educational Research Journal, 1966, 3, pp. 125-130.

Scott, Dennis L. and Webb, Leland F. The effects of spacing and handwritten or typed tests upon achievement of students in algebra and grometry classes. School Science and Mathematics, 1979, 79 (8), pp. 663-669.

Shared Evaluation Handbook, Division of Instruction, Department of Education, Government of Newfoundland and Labrador, 1980.

Small, Dwain E., Holtan, Boyd D. and Davis, Edward J. A study of two methods of checking homework in a high school geometry class. The Mathematics Teacher, 1967, 60 (2), pp. 149-152.

Starch, Daniel and Elliott, Edward C. Reliability of the Grading of High School Work in English. School Review, 1912, 20, pp. 442-457.

________. Reliability of Grading Work in History. School Review, 1913, 21, pp. 676-681.

________. Reliability of Grading Work in Mathematics. School Review, 1913, 21, pp. 254-259.

Suydam, Marilyn N. Evaluation in the Mathematics Classroom. The ERIC science, mathematics and environmental education and the center for science and mathematics education. The Ohio State University, 1974.

Taylor, P.H. A study of the effects of instructions in multiple-choice mathematics test. British Journal of Educational Psychology, 1966, 36 (1), pp. 1-6.

Terwilliger, James S. Assigning Grades to Students. Scott, Foreman and Company, 1971.

Thorndike, Robert L. Essentials of Educational Measurement. Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1972.

Thorndike, R.L. and Hagan, E. Measurement and Evaluation in Psychology and Education. John Wiley and Sons, New York, 1977.

White, M.A. and Boehm A. Child world of learning: written workload of pupils. Psychology in the Schools, 1967, 6, pp. 70-73.

APPENDIX A

The Instrument

ITEM 1

QUESTION: Of three consecutive integers, the sum of the first two plus four times the third is 99. Set up and solve an equation to find the integers.

SOLUTION:

Let X = the first
X + 1 = the second
X + 2 = the third

$$X + X + 1 + 4(X + 2) = 99$$
$$2X + 1 + 4X + 2 = 99$$
$$6X + 3 = 99$$
$$6X = 99 - 3$$
$$6X = 96$$
$$X = 96/6$$
$$X = 16$$

The integers are 16, 17, and 18.

7

Please specify why the student lost points.

point value out of 10 [ ]

____

____

____

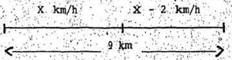
____

Correct Answer = (15, 16, 17)

ITEM 2

QUESTION: Bill and Jane left points 9 km apart to walk toward each other. Jane walked 2 km per hour slower than Bill. They met in 3 hours. Set up and solve an equation to find their speed.

SOLUTION: Let X represent Bill's speed and X - 2 represent Jane's speed.



$$X + 3 + X - 2 + 3 = 9$$
$$2X + 3x + 1 = 9$$
$$2X + 4 = 9$$
$$2X = 9 - 4$$
$$2X = 5$$
$$X = 5/2 \qquad \text{If } X = 2.5$$
$$X = 2.5 \qquad X - 2 = .5$$

Bill's speed was 2.5 km per hour and Jane's was .5 km per hour.

Please specify why the student lost points.

point value out of 10

Correct Solution = (2.5, .5)

ITEM 3

QUESTION: This year 47 students turned out for the swimming team. This is 11 more than 2/3 the number last year. Set up and solve an equation to find the number that turned out last year.

SOLUTION: Let X = the number of students last year.

$$2/3 \cdot X + 11 = 47$$
$$2/3 \cdot X + 11 - 11 = 47 - 11$$
$$2/3 \cdot X + 0 = 36$$
$$2/3 \cdot X = 36$$
$$3/2 \cdot 2/3 \cdot X = 3/2 \cdot 36$$
$$X = 64$$

There were 64 students last year.

Please specify why the student lost points.

point value out of 10 ☐

Correct Solution = (54)

ITEM 4

QUESTION: A collection of stamps is worth \$2.80. There is one less 5¢ stamps and two more 20¢ stamps than 10¢ stamps. Set up and solve an equation to find the number of 10¢ stamps.

SOLUTION: Let X = the number of 10¢ stamps.
then X - 1 = the number of 5¢ stamps.
and X + 2 = the number of 20¢ stamps.

$$10X + 5(X - 1) + 20(X + 2) = 280$$
$$10X + 5X - 5 + 20X + 40 = 280$$
$$15X - 5 + 20X + 40 = 280$$
$$35X + 35 = 280$$
$$70X = 280$$
$$X = 4$$

There are four 10¢ stamps.

Please specify why the student lost points.

____

____

____

____

point value out of 10 ☐

Correct Solution = (7)

ITEM 5

QUESTION: Of two digits, the second is three less than the first. If the first is increased by 2 and the second is doubled their sum will be 8. Set up and solve an equation to find the first digit.

SOLUTION: Let X = the first digit
and X - 3 = the second digit

$$X + 2 + 2(X - 3) = 8$$
$$X + 4(X - 3) = 8$$
$$X + 4X - 12 = 8$$
$$5X - 12 = 8$$
$$5X = 8 + 12$$
$$5X = 20$$
$$X = 20/5$$
$$X = 4$$

The first digit is 4.

Please specify why the student lost points.

point value out of 10 ☐

____

____

____

____

Correct Solution = (4)

ITEM 6

QUESTION: The length of a rectangle is 3 cm shorter than twice the width. The area is 135 cm². Set up and solve an equation to find the length and width.

SOLUTION:

Let X = the width
and 2X - 3 = the length

$$X + 2X - 3 = 135$$
$$3X - 3 = 135$$
$$3X = 135 + 3$$
$$X = 138$$
$$X = 138/3$$
$$X = 46$$

$$\text{Length} = 2X - 3$$
$$= 2(46) - 3$$
$$= 92 - 3$$
$$= 89$$

The width is 46 cm and the length is 89 cm.

Please specify why the student lost points.

____

____

____

____

point value out of 10 ☐

Correct Solution = (9, 15)

ITEM 7

QUESTION: For a theatre production, student tickets cost \$0.75 and adult tickets cost \$1.25. Altogether 260 tickets were sold. The total proceeds were \$249. Set up and solve an equation to find how many tickets of each kind were sold.

SOLUTION: Let X = the number of student tickets and 260 − X = the number of adult tickets

$$.75X + 1.25(260 - X) = 249$$
$$.75X + 335 - 1.25X = 249$$
$$-.50X = 249 - 335$$
$$-.50X = -86$$
$$X = -86/-.50$$
$$X = 172$$

Adult tickets
= 260 − X
= 260 − 172
= 88

There were 172 student and 88 adult tickets sold.

Please specify why the student lost points.

point value out of 10 ☐

____

____

____

____

Correct Solution = (152, 108)

ITEM 8

QUESTION: A 13 cm board is to be divided into three pieces. The second piece will be 3 cm longer than the first, and the third piece will be twice the second. Set up and solve an equation to find the length of the shortest piece.

SOLUTION: Let X = the shortest piece.
and X + 3 = the second piece.
and 2(X + 3) = the third piece.

$$X + X + 3 + 2(X + 3) = 13$$
$$X + X + 3 + 2X + 6 = 13$$
$$2X + 3 + 2X + 6 = 13$$
$$4X + 9 = 13$$
$$13X = 13$$
$$X = 1$$

The shortest piece is 1 cm.

Please specify why the student lost points.

point value out of 10

Correct Solution = (1)

APPENDIX B

Letter Attached to Instrument

P.O. Box 35
Education Building, MUN
St. John's, Newfoundland
A1B 3X8
April 22, 1982

Dear Colleague,

This survey is being conducted to determine the grading practices of teachers in assigning partial credit to students in Mathematics. In the following questionnaire you will find word problems with typical student solutions supplied. Many of the solutions were taken from actual student work.

You will find that none of the solutions are completely correct. Please grade each of the questions as you would normally do for any chapter or unit test. Each question should be graded out of 10 points and the grade placed in the space provided. The correct solution is indicated at the bottom of each page.

After grading the student's answer, please specify in the space provided, why the grade was assigned. Students were not required to show the checking of their solution so no marks should be deducted for this.

Thank you very much for your time.

Sincerely,

Patricia J. Maxwell

APPENDIX C

Cover Letter to Principal

P.O. Box 35
Education Building, MUN
St. John's, Newfoundland
A1B 3X8
April 22, 1982

Dear Principal,

As part of the requirements for a Master of Education degree in Curriculum and Instruction, I am conducting a survey to determine the grading practices of teachers in assigning partial credit to students in Mathematics.

I would appreciate it very much if you would assist me in my research by asking your teacher(s) of the Level I (Grade 10) Academic Mathematics course to complete the enclosed questionnaires. Should you have more teachers of the subject than the number of questionnaires enclosed, please select the teachers who teach the greater number of Level I Academic Math classes. If you receive more questionnaires than there are teachers of the subject in your school, please enclose any blanks with your returns.

While I realize that this is a somewhat hectic time of year for all educators, I would very much appreciate your assistance and that of your Mathematics teacher(s) in this matter. Please be assured that all replies will be confidential, in that no one school or teacher can be identified from the study.

When the questionnaires are completed, please detach introductory letter, fold questionnaires, and place them in the self-addressed, stamped envelope which is enclosed.

Thank you so much for your cooperation.

Sincerely,

Patricia J. Maxwell

APPENDIX D

Reminder Letter

P.O. Box 35
Education Building, MUN
St. John's, Newfoundland
A1B 3X8
May 4, 1982

Dear Principal,

Approximately two weeks ago a questionnaire concerning teacher grading practices in Mathematics was sent to your school. Since the results of this survey are vital to the completion of my M.Ed. degree, I would very much appreciate your efforts in supplying me with a speedy return. If, however, you have already mailed your returns please disregard this letter.

Thank you very much for your time.

Sincerely,

Patricia J. Maxwell