DOCUMENT RESUME

ED 300 443 TM 012 488

AUTHOR	Traub, R.; And Others
TITLE	Teacher Assessment Practices in a Senior High School Mathematics Course. Final Report.
INSTITUTION	Ontario Inst. for Studies in Education, Toronto.
PUB DATE	Jun 88
NOTE	133p.; Some appendices are marginally legible.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	*Calculus; *Classroom Techniques; College Bound Students; *College Preparation; Educational Assessment; Foreign Countries; *Grading; Mathematics Tests; Postsecondary Education; Secondary Education; Secondary School Mathematics; Secondary School Students; *Secondary School Teachers; *Student Evaluation; Testing
IDENTIFIERS	*Ontario

ABSTRACT

The relationship between teaching and assessment in Ontario (Canada) was studied via analysis of classroom practices, content of assignments and tests, grading practices, and responses to a common marking task of 17 Grade 13 Calculus teachers. Research on means of student assessment procedures for Grade 13 examinations was initiated in 1986. Since then, student assessment has been conducted by experienced teachers. This study focused on whether teachers were assigning course marks that were comparable in terms of curricular emphasis, the kinds of evidence collected about student achievement, criteria used by teachers to assess evidence, and marking standards. The grade 13 (pre-university) calculus course is successfully completed by 30,000 Ontario students each year. After a pilot study during the fall semester of the 1986-87 school year, a spring semester main study was conducted, followed by a study of marking standards during the fall of 1987-88. Activity logs, supplemental teaching materials, assessment instruments, marked examination papers, and final mark records were reviewed. Examinations and term tests were the main determinants of student grades. Results show: wide variations across teachers in allocation of classroom time to various activities; substantial differences in content emphasis; substantial agreement among the teachers as to the relative importance of examination questions and relative quality of 20 sample student papers; and substantial disagreement among the teachers as to the absolute quality of the student papers. (TJH)

Teacher Assessment Practices in a Senior High School Mathematics Course

1987-88 Transfer Grant
Project 52-1028
The Ontario Institute for Studies in Education

R. Traub and P. Nagy
Principal Investigators

and

K. MacRury and R. Klaiman
Research Officers

Final Report[1]
June, 1988

[1]We wish to thank the 17 teachers who gave so generously of their time to make this study possible. We also want to thank the members of the advisory committee for their assistance.

Table of Contents

Executive Summary

The study was motivated, at least in part, by the widespread concern in the province of Ontario over the apparent lack of consistency in the meaning of school-leaving marks. It has been claimed, for example, that students in first-year college and university classes differ markedly in knowledge of mathematics, and that these differences are not reflected by corresponding differences in high school grades in mathematics. Provincial Grade 13 examinations were discontinued in 1968, and since then teachers have been responsible for student assessment.

The aim of the project was to describe how student assessment was being conducted by experienced teachers who were teaching the same senior high school mathematics course, and to determine whether these teachers were assigning course marks that were comparable in terms of curricular emphasis, the kinds of evidence collected about student achievement, the criteria used by teachers to assess this evidence, and marking standards.

The course selected for study was Grade 13 (pre-university) Calculus. One of three Grade 13 mathematics courses, it is successfully completed by about 30,000 Ontario students each year. A credit in the Calculus course is a prerequisite for admission to several undergraduate faculties (e.g., engineering, science) at Ontario universities.

To be eligible to participate in the study, Calculus teachers were required to meet the following minimum qualifications: (i) an undergraduate degree in mathematics, (ii) five years experience teaching senior mathematics, and (iii) three years experience teaching calculus.

Phase 1, the pilot study, was conducted with a group of teachers in the Fall semester of the 1986/87 school year for the purpose of refining the instrumentation and procedures. For Phases 2 and 3, a second group of twenty teachers from different semestered schools in fourteen school boards was recruited. Seventeen of the twenty teachers completed their committment to the project. With one exception, the seventeen teachers, from thirteen large Southern Ontario school districts, taught in medium to large schools; one school had 300 students, twelve had between 900 and 1500 students, and four had between 1800 and 2100.

Phase 2 was the main study. For the duration of the Spring 1986/87 semester, each participating teacher (i) maintained a daily log of activities for the Calculus class, and provided (ii) copies of all supplemental teaching materials and assessment instruments, (iii) a class set of marked examination papers, and (iv) the final record of marks for the class, including marks for all assignments, tests and examinations used in deriving students' final grades.

Phase 3, a study of marking standards, was conducted during the Fall 1987/88 school semester. The same group of teachers developed a marking scheme, marked, and then returned a set of calculus test papers provided by the project team.

From all parts of the study, a picture emerged of variation from classroom to classroom. The analysis of use of classroom time showed: (a) total class time required for the credit ranged from 96 to 114 hours (median 105); (b) time spent on testing ranged from 9 to 17.2 hours (median 12.5); (c) percentage time spent on teacher-centred direct instruction ranged from 17% to 52% (median 26%); (d) percentage time devoted to seat-work ranged from 8% to 46% (median 26%); and (e) percentage time devoted to review ranged from 4% to 14% (median 11).

Results of the examination of assignment content and testing content showed: (a) teachers assigned between 450 and 1600 questions during the term (median 1040); (b) the proportion of assignment questions on basic algorithms, ranged from 41% to 76% (median 56%); (c) on average, teachers awarded 18% fewer marks to testing basic algorithm practice than their assignments indicated —

across teachers, this figure varied from 6% to 30%; and (d) with care, an exemption policy can allow comparison between exempted and un-exempted students — this is not the case for the policy under which a student is allowed to drop the worst test result of the term.

The analysis of grading policies and practices showed: (a) grading criteria consisted almost entirely of two forms of testing, term tests and examinations; (b) teachers gave as few as 3 and as many as 13 term tests; (c) final grades were lower than midterm grades by from 0 to 15 marks (median 6); and (d) final examinations had a standard deviation on average 50% higher than did term marks (this increased to 100% higher for schools with exemption policies).

Our analysis of teacher responses to a common marking task showed: (a) teachers varied greatly in the number of marks allowed in marking schemes for different questions, but they agreed substantially on the relative importance of different items; (b) teachers agreed very well on the relative ranking of the student exams, but they disagreed profoundly in the grades awarded: the hardest marker awarded no honours and failed 10 of the 20 papers, while the easiest marker awarded 7 honours and gave no failures; (c) standards expected by different teachers varied more for higher-level thinking questions than for basic skills questions.

This study, intensive and involving few teachers, was not designed to find generalizable patterns across the different aspects of the study. More extensive work, involving more teachers, is required to uncover such patterns.

Introduction

The purpose of this report is to examine possible reasons for variations in the meaning of grades assigned by different teachers in the same course. In Ontario, the high school curriculum is constrained by provincial guidelines, which specify minimum course content and length within a broadly defined educational philosophy. Local school boards are responsible for implementation of the guidelines in their schools, with consequent differences in topic emphasis, grading methods, and performance expected for a given mark. These differences were the subject of the present study.

The specific purpose of this two-year investigation was to examine the relationship between teaching and assessment in Grade 13 Calculus. To this end, we obtained and examined data of several types from a sample of 17 teachers of calculus. These teachers:

— completed daily logs of time spent in different classroom activities, using a set of categories provided by the research team;

— provided, on the same log form, lists of the homework and seat-work assigned each day;

— reported the criteria used to arrive at student grades for the course, including tests, quizzes, examinations, and other factors (e.g., participation, attendance), along with the relative weights of each;

— marked, using a marking scheme of their own design, a common set of 20 final examination papers obtained from a class not involved in the study.

With this information, we were able to examine patterns in the use of classroom time, variations in content taught, discrepancies in the match between what was taught and what was tested, differences in amount and kind of testing, and differences in expectations for quality of student response.

Related Literature

Research on classroom assessment is a relatively recent phenomenon, confined largely to the past decade. Before this, research focussed on the use of standardized test results by classroom teachers. Reviews of this research indicated that, although teachers use standardized test results to confirm their own judgements of student achievement, little is known about how they make the judgements in the first place (Lazar-Morris, Polin, Moy & Burry, 1980; Rudner, 1984). To illustrate the lack of attention to teacher-developed assessment by the educational measurement community, several authors (Stiggins, Conklin & Bridgeford, 1986; Terwilliger, 1987) have cited a special 1983 issue of the *Journal of Educational Measurement*, which was devoted to linking testing and instruction, but which specifically excluded research on teacher-made tests (Burstein, 1983).

The questionnaire survey is the most frequently reported method of collecting data on teacher assessment practices and training in measurement and student evaluation. Several surveys have been reported within the last five years. The evidence collected in these surveys supports the following conclusions:

— Teacher-developed tests and classroom observations are the most common methods for assessing students (Herman & Dorr-Bremme, 1983).

— Variations in assessment practices exist across grade levels. For example, elementary teachers tend to favour observations and secondary teachers testing (Stiggins & Bridgeford, 1985). Elementary teachers use more diagnostic tests and completion and matching items; secondary teachers use more performance and competency tests, with short answer items used in middle school and objective and essay items used increasingly in high school (Green & Stager, 1986).

— Variations in assessment practices exist across subjects, particulary at the secondary level. Mathematics is heavily test-oriented compared to other subjects (King et al, 1976; Wahlstrom & Danley, 1976). Mathematics test items tend to be of the short answer format, and English test items tend to require more extended responses and the exercise of more subjectivity in grading (Green & Stager, 1986).

- Terwilliger (1987) found similar subject differences in his comparative study of assessment practices of teachers in Minnesota and England, but he also concluded that the differences were a function of both the subject and the educational system. For example, Terwilliger attributed the Minnesota teachers' preference for objective types of tests to the attention given to objective testing in American textbooks on educational measurement; similarly, he considered the importance placed by the teachers in England on such non-cognitive variables as attendance and effort to be a reflection of the traditions and values of the English system.
- Teachers lack training in both the teaching and assessing of higher-order thinking skills (Haertel, 1986; Stiggins, 1988).
- The measurement community needs to learn more about how assessments of student achievement are conducted within the classroom by using such research methods as classroom observation (Stiggins & Bridgeford, 1985; Stiggins, Conklin & Bridgeford, 1986; Anderson, 1987).

A variant on the teacher survey approach is the analysis of teacher-developed tests and examinations (Fleming & Chambers, 1983; Graham, 1984; Alexander, 1987; Marso & Pigge, 1988). This analysis has focussed on such characteristics as item format, test reproduction and cognitive level of the items. The main conclusion of this research is that most teacher-developed test and examination questions are pitted at the knowledge level of Bloom's taxonomy (Bloom, 1956). This work is flawed, however, in that the test and examinations have not been considered in relation to the other forms of assessment used in the classroom.

There are few examples of studies in which methodological alternatives to the survey approach have been employed. One such example is Stiggins' (1988) classroom observation study in which teachers were observed in their classrooms for one day, with oral questions being recorded and analyzed to determine cognitive level. A limitation of this study is the length of the observation period. One day is not necessarily sufficient to (i) yield generalizable information about a teacher's questioning behaviour or (ii) document the classroom context at the level of detail needed to discern possible explanatory factors. The financial cost of a more extended study is, unfortunately, often prohibited.

In another study in which an alternative (to the survey) method is used, Wilson (pending) collected all the formal evaluation instruments used throughout 36 course offerings in one Ontario high school. Wilson's findings included the following:

— On the frequency of evaluation, over the 36 courses there was an average of 107 hours of instruction and a type of formal evaluation activity occurring every ten hours. Frequency appeared to be a function of teacher preference and not a function of subject area or level of course.

— On the purpose of evaluation, two goals were expressed for almost every instrument: to have students practice or apply learning and to generate marks for reporting purposes.

— To accommodate both instructional purposes (more objectives-based needs) and administrative purposes (more norm-referenced needs), Wilson found that teachers used more frequent but less heavily weighted evaluations in the first half of the semester, and, "[a]s the time for mark reporting drew near, more heavily weighted but less frequent tests and examinations were used to bring the total mark distribution into line." (p. 9)

The aforementioned studies have for the most part provided evidence on which to base comparisons of the assessments made for different subjects and grades. The typical, and not unexpected, finding is that differences in assessment practices exist across the curriculum. With the exception of Graham (1984) and Alexander (1987), however, the practices within a subject and grade have not received much attention; this type of comparison is attempted in the present study.

Project History

The project began in the spring of 1986. One of the first tasks was to choose the mathematics course to be studied. Due regard had to be given, in

The fall of 1986 was taken up with the development, pilot testing, and revision of the form for the daily log that teachers were to keep throughout the course. In addition, at this time, a two-week pilot study was conducted to assure ourselves that the data collection procedures were both feasible and palatable to teachers. The main body of data was collected during the Spring 1987 semester. Responses to a common marking task were obtained in the fall of the same year.

Outline of Report

First, we examine differences in teacher use of classroom time. Given that we relied on teacher self-reports, our analysis of classroom time is somewhat circumscribed. It provides context for the rest of the report, and allows additional characterization of teachers, whose teaching and assessment practices proved interesting in other respects. Second, we analyze the content of the teaching and the testing, and consider how well they were matched. Third, we examine differences in testing policy and practice. This examination includes how different components of the final grade were weighted, and how student data were gathered to set the final grade. Finally, we examine the grades awarded by the teachers to a common set of examination papers. We close with a summary, and suggest several questions to be addressed in future work.

making this choice, to the fact that a new secondary mathematics curriculum was being introduced in Ontario, and the transition was being made, at different times by different schools, from Grade 13 courses to Ontario Academic Courses (OACs). It was also important that the results of the study be relevant. A comparison of Ministry Guidelines for Grade 13 and OAC mathematics courses (Ontario Ministry of Education, 1972, 1985) revealed that the only courses that overlapped substantially were Grade 13 and OAC Calculus. Choice of calculus also meant that the results of the survey of calculus examinations by Alexander (1986) would be available to complement and inform the results obtained in this study.

The decision to focus on calculus was taken after consulting with the advisory committee that had been formed for the project. The advisory committee consisted of the following five individuals:

- Dr. David Alexander, Faculty of Education, University of Toronto, and The Ontario Ministry of Education.
- Dr. Edward Barbeau, Department of Mathematics, University of Toronto
- Dr. Gila Hanna, Department of Measurement, Evaluation and Computer Applications, The Ontario Institute for Studies in Education.
- Mr. George McNabb, Mathematics Teacher, Sudbury Board of Education, representing The Ontario Association of Mathematics Educators.
- Mr. John Scott, Mathematics Consultant, Toronto Board of Education.

The design for the study called for twenty teachers to be recruited as participants in the main data collection phases of the project. Cost restricted the choice of teachers to those working in the urban core of Southern Ontario. The twenty recruits taught in 20 different schools and represented 14 different boards. (The teachers were guaranteed anonymity, so their names must remain confidential.) Three eventually dropped out for reasons unrelated to their participation in the project. All were qualified and experienced teachers of Grade 13 Calculus. Each held an undergraduate degree in mathematics, had at least five years experience teaching senior mathematics, and had taught the calculus course at least three times.

Classroom Practice

Each of the 17 teachers in the study maintained a daily log for one calculus class for an entire semester. The purpose of the logs was to obtain a description, for each class period, of (i) instructional activities, including the topic(s) covered each day, the time spent on teacher and student-centered activities, the examples used, homework assigned and problems encountered, and (ii) assessment activities, including time used for assessing student achievement, assessment devices (including grading schemes), and marks assigned.

The logs were maintained on a specially prepared form. This form provided space for teachers to state the date and objective(s) of the class, indicate the sequence of class activities and specify the time (in minutes) spent on each activity, the focus of the activity and the major player (teacher or student). (Pilot work on several versions of a log led us to one in which the directions for completing the log were general enough to allow the teachers freedom to describe classroom practices in their own terms.) A copy of the log is included as Appendix A. Inasmuch as teacher logs are self reports and inasmuch as corroboratory observation data were not obtained, we are uncertain of the extent to which the teachers were able to provide accurate reports of classroom activities and of the time allocated to each activity.

In their daily logs, most teachers provided descriptions of classroom practice that preserved the order, duration and focus of the teaching/learning activities of each class period, and indicated how class time was used to teach the topics and subtopics. Four teachers, however, concentrated more on topics than on activities, and a fifth delegated the task of completing the log, assigning it on a rotational basis among students. The logs of the four did not include sufficient detail to identify activities or permit the amount of time used for each activity to be ascertained, and the logs of the fifth were uneven in format.

Consequently, the logs of these five teachers were excluded from further consideration in this part of the study, leaving the data for 12 teachers to be considered.

Results

Basic descriptive data on school and class size, along with total teaching time and number of tests given, are reported in Table 1. Note that, with one exception, the schools are moderate to large in size.

****

Insert Table 1 about here

****

From 80 to 109 logs (Mean = 86) were obtained from each of the 12 teachers for whom a detailed analysis of logs was made. The length of the class periods ranged from 10 minutes, in one instance, to 85 minutes (Mean = 66 minutes). A comparison of logged time with the school timetable revealed a difference between scheduled time and actual time of as much as 10%. Reasons given for the loss of scheduled time included class cancellations for school assemblies, special events such as career days and graduation activities, religious holidays, and teacher illness. Similarly, shortened periods often resulted from unplanned events, such as special assemblies and fire drills, although not all teachers provided reasons for abbreviated periods.

Six categories and two subcategories of activities were defined from terms used in the logs to describe class activities:

1. Administration: taking attendance, making announcements, planning presentation groups and similar non-instructional uses of classroom time.
2. Direct instruction: teacher-centered activities, such as presentations, demonstrations, discussions and lectures focussing on new material.
3. Student practice: student-centered activities, such as seat-work and board-work pertaining to new material (including handouts, assignments and orally presented problems considered in class), with opportunity for individualized

instruction. A subcategory, the review exercise, includes handouts and textbook problems assigned for review purposes under similar instructional conditions.

4. Homework: tasks assigned for independent completion, either in class time or outside, and later considered in a class discussion.
5. Review: class time used (i) to cover previously learned material, including prerequisite knowledge acquired in other courses (e.g., algebra), content previously covered in the course, and (ii) to prepare for tests and exams. A subcategory, test review, included class time used for marking and/or discussing answers to tests, quizzes and exams.
6. Assessment: quizzes, class tests and exams administered in class time.

Analysis of the logs focussed on the time allocated to different instructional activities. Because the length of classes varied, the time devoted to each category of activity was calculated as a percentage of logged time. The percentage of class time allocated in each of the 12 classes to each of the six activities is shown in Table 2. On average over the 12 teachers, the activities in decreasing order of percentage of time are: student practice (29.1%), direct instruction (27.9%), homework (20.4%), assessment (11.5%), review (9.6%) and administration (1.1%). Comparing the mean percentage of class time given to student practice and homework as opposed to direct instruction, one sees that 50% of class time was devoted to student-centred activities, while teacher-centred direct instruction amounted for 28% of the time.

****

Insert Table 2 about here

****

Table 2 also shows the distribution of activities for each of the 12 teachers. Every category of activity was reported by each teacher, with the exception of Teacher 4, who reported no use of class time for administration or for review exercises. The teachers varied considerably in the amount of time they allocated to the different activities. The time for direct instruction, for example, varied from 50% (Teacher 1) to less than 20% (Teachers 11 and 17).

Teacher 11 used more time for homework than any other teacher, but devoted a similar proportion of time to seat-work as Teacher 1. Teacher 16, in contrast, used the most time, almost 50%, for seat-work.

In the search for patterns in the data reported in Table 2, coefficients of correlation were computed (over teachers, between columns). The results showed a significant, negative correlation between total logged time and the percentage of time spent on direct instruction ($r = -0.66$, $p < .01$). The possibility that teachers with greater amounts of class time organize to do less direct teaching is also suggested by the findings that percentage of time for direct instruction correlated negatively with percentage of time for homework ($r = -0.63$, $p < .01$) and student practice ($r = -0.37$, ns), whereas percentages of time for homework and for student practice are each positively but not significantly correlated with total time.

The largest percentage of time, overall, was devoted to student practice (see Table 2). Highly significant negative correlations between percentage of time for student practice and percentages of time for review ($r = -0.85$, $p < .01$) and for assessment ($r = -0.75$, $p < .01$) suggest that teachers who place relatively high emphasis on practice in their teaching of calculus place a relatively low emphasis on review and assessment activities.

Content of Assignments and Tests

The purpose of this part of the study was to examine (a) the content of assignments, (b) the content of tests, the results of which were used to arrive at a final grade, and (c) relationships between the two. The motivation for this part of the study stems from concern for the comparability of grades awarded by different teachers: to what extent are the grades of different teachers based on the same content coverage? The data come from the daily logs of homework and seat-work assignments, and details of quizzes, tests and exams, associated marking schemes, and the relative weights of factors that determine the final course grade. This part of the report is focused on content. No distinction is made between different methods of arriving at marks (i.e., quizzes, term tests, or exams), except for making use of their relative weights as reported by the teachers.

A comment on the context for this study is appropriate, lest our results be taken as implicitly critical of the teachers who cooperated in its conduct. The Guideline for the Ontario Grade 13 Calculus Course (Ontario Ministry of Education, 1972) mandates broad content areas, but not relative importance. Thus, we are not in. .tigating whether some teachers exercise better or poorer judgement as to what should be in the curriculum. Instead, we are simply investigating the different judgements as to content that have been made by qualified and experienced teachers. As well, some of our data reveal considerable mismatch between relative emphasis of topics in assignments as compared to tests. Here again, this mismatch is not necessarily indicative of bad pedagogy. Our analysis ignores the issue of whether one topic is logically prerequisite to another, so, for example, it may be entirely appropriate to teach Topics X and Y, yet test only Topic Y. The latter may, in fact, be legitimately viewed as Topic X at a higher conceptual level. Our perspective, then, is that we are examining differences in coverage, and not passing judgement on such differences.

This section of the report begins with a discussion of problems with the data. This is followed by a description of the development of a system for categorizing the content of the calculus course, and an application of this system to the questions assigned by the teachers. Then the system is applied to questions from tests, quizzes, and exams, and the similarities and differences between testing content and assignment content are examined. In a side analysis, two common grade assignment practices are examined for their effect on the basis of the final grade: allowing students to drop their least successful test(s) in the calculation of a final grade, and exempting students with sufficiently high term performance from the final exam. Finally, the analysis is repeated at a finer level of detail for one topic within the calculus course, and for a subset of the 17 teachers.

Problems With The Data

There are four problems with the data, each of which impinges on the accuracy and completeness of what is reported. First, despite detailed instructions to teachers and frequent contact during data collection, some major differences exist in the quality of information obtained from different teachers. Of primary importance is a lack of detail on expectations for assignment completion. For example, the ratio of the largest to smallest number of questions assigned over the semester by a teacher was about 3.5. We do not know, however, if a given teacher had the attitude "try enough of these to make sure you understand the topic" or "I expect you to do all of these." Although we know the homework and seat-work review practices of the teachers — most discussed only those questions students reported difficulty with — there remains an element of doubt as to whether number of questions assigned is a direct measure of amount of effort required. Second, a clearly defined natural unit of analysis is lacking. Content emphasis of tests was analyzed on the basis of percentage of marks allocated to each content category represented in the items on the test, multiplied by the relative weight of the test towards the final course grade. For practical reasons — sheer volume of data and lack of teacher-provided weights — the unit of measurement for the analysis of assignments is the question. Neither marks nor questions are of uniform size. For example, a drill-and-practice question on the differentiation algorithm will typically take less time than a maximum/minimum problem. We have tried to minimize the effect of this problem by avoiding direct comparisons of numbers of questions across topics, and focusing instead on relative (across teachers) importance.

The third problem with the data is missing information. Five teachers on a total of 14 occasions assigned an indefinite number of questions from a section of text. These missing values were estimated from marginal information on relative category popularity and relative number of questions assigned by the teacher.

Fourth, variations in the methods used to arrive at grades prevented simple and direct comparison across teachers. To facilitate these comparisons, several approximations were used:

— Ten of 17 teachers used tests and exams for less than 100% of their final grade; up to 20% of the final mark was determined by participation and assignments. In these cases, we ignored the participation and assignment categories and readjusted the weights for tests and exams so that they totalled 100%.
— Ten of 17 teachers used bonus questions, making the available number of marks for a test or exam more than 100%. Here, we readjusted the teacher-provided weights proportionally downward to 100%.
— Four teachers had a policy of exemption from the final exam for those students whose term marks were sufficiently high. In the main body of the data, grading weights are reported for the un-exempted students, and the effect of the exemption policy is reported in a side analysis.
— Five teachers followed policies that allowed students to drop one or more of their poorer test results. These policies were ignored in the main analysis, but the effect of one example of this policy is reported in another side analysis.
— One teacher allowed differential weighting between final exam and term mark, favoring the higher of the two (65-35 versus 35-65). Since only 4 of

21 students did better on the final than the term, and then just by 1% or 2%, only the majority situation is reported.

In short, we have used several approximations intended to smooth out idiosyncracies that have been, to some extent, the subject of other parts of this project. The reward is comparison of individuals on a common basis; the price is that this comparison is slightly removed from the reality of classroom grading.

The Category System

The first step in the analysis was development of a categorization scheme for the content of the course. This process began with an examination of the 1972 Grade 13 Calculus Guideline (Ontario Ministry of Education, 1972) and the contents of the two Ministry approved texts for the course. Refinements, corrections, and additions to the category system continued during much of the analysis. All questions from the two texts were categorized by members of the research team, with advice from the members of an advisory panel of mathematics educators that was formed for the project. When a satisfactory version of the category system had been produced, we enlisted the services of two students, then in the graduating term of a B.Sc./B.Ed. program in mathematics education. Both students had recent experience in practice teaching the Calculus course, and were recommended by a faculty member familiar with their abilities. These students first reviewed and revised the category system while applying it to the textbook questions, and then applied the final version of the system to questions on teacher-produced handouts, tests and exams.

The final version of the category system included 126 topics. In applying this scheme to common material, the two students achieved an inter-rater agreement was 88%. To simplify reporting of results, the 126-topic scheme was collapsed into 14 larger categories; at this level, inter-rater agreement was 97%. The more detailed 126-topic system is reported in Appendix B; the 14-Category system and a 6-Group version are reported in Table 3. The hierarchical relationship of the three category systems is described in Appendix C. We will be consistent in referring to the 14 content areas as Content Categories and the six as Content Groups.

****

Insert Table 3 about here

****

The six content groups reflect major divisions and emphases to be expected in the curriculum. Group I contains the basic core skills of calculus, while Group II includes topics associated with the theoretical basis of the core skills. The topics in Groups III and IV represent the graphical application of basic skills, and have been separated because they are substantially different in emphasis and difficulty. The topics in Group III reflect understanding of differentiation, while those in Group IV refer to a relatively difficult application. Group V topics involve the application of basic skills to situational problems, and Group VI topics are optional — they might be placed elsewhere than in a calculus course.

Results

Analysis of Assignments

Table 4 is a record of the number of questions assigned by each teacher that were sorted into the 14 content categories. The variations among teachers in total number of questions assigned was wide. The median of the distribution over teachers of number of questions assigned was 1037; the difference between the 25th and 75th percentiles of the distribution (Q1-Q3 range) was 482. Exploratory analysis sheds some light on the source of this huge variation. Stem-and-leaf displays[1] (medians removed, unit = 10) are provided below for the total number of questions assigned and the total number excluding the questions

[1] A stem-and-leaf display is a kind of frequency distribution. Associating a leaf with its stem reproduces an observation tabulated in the display. In the display on page 16, under the heading All Items, the leaf 8 with the stem -5 yields the observation -58, but since the unit is 10 in this diagram, the observation tabulated was -580. The teacher to which this observation applies assigned 580 fewer questions than the median number of 1037 questions.

assigned for the four topics in Content Group I. Note that the latter display contains less variability than the former, indicating that a substantial portion of the variation in total number of questions assigned is due to variation in number of questions assigned at the skill level.

All Items		With Skill (Group I) Items Removed	
Stem	Leaf	Stem	Leaf
-5	87	-2	60
-4		-1	510
-3		-0	850
-2	0	0	03599
-1	82	1	6
-0	763	2	8
0	04	3	79
1			
2	3		
3	0055		
4	6		
5	8		

The Q1-Q3 range with skill level questions removed is only 238 (median 425). It might be argued that skill-level questions on differentiation (Category 3 of Group I) form the core of a course in calculus. Examination of Category 3 numbers in Table 12 reveals that the two teachers (4 and 6) with lowest total numbers of questions assigned fewer than 200 differentiation skill questions, whereas those teachers with the highest totals (3, 5, and 8) assigned more than 500 such questions.

****

Insert Table 4 about here

****

The numbers in Table 4 were summed over content categories within content groups and then converted to percentages of (row) totals for each teacher. The results are reported in Table 5. (Percentage values for the 14 content categories are given in Appendix D.) As can be seen, the pattern of variation in percentages over content groups is similar for each teacher, with most emphasis given to basic skills (Content Group I) and least to proofs (Group II). Also striking is the variation over teachers in the emphasis placed on optional topics (Group VI).

****

Insert Table 5 about here

****

Inspection of pair-wise plots and correlations of the columns in Table 5 and Appendix D reveals several trends that can be used to characterize teaching emphasis. First, the percentage of questions on the basic skills (Content Group I) range from 40.6% (Teacher 10) to 75.5% (Teacher 8). Emphasis of basic skills is strongly associated with deemphasis of almost everything else. Second, Content Groups III, IV and V are somewhat interrelated; emphases on graphing and situational problems tend to go together, although teachers high in one and not the other can be identified. Third, it can be seen that little attention was paid to questions involving proofs and first principles, although increased emphasis on these issues is mandated in the new Ontario curricula. Only 4 of 17 teachers assigned more than 1% of their questions from within this group, and none had more than 3%. Stress on proofs and first principles, however, where it occurs, is associated with an emphasis on integration skills (Category 4 of Content Group I) and a deemphasis on integration graphing (Content Group IV).

The following is offered as a tentative categorization of teaching emphasis, with teachers in a subjectively determined order (bracketed numbers refer to teachers whose placement is uncertain):

High on basic skills (Group I): 8, 5, 12, 7, 15, 3, (6).
High on optional topics (Group VI): 1, 9, 16.
High on situational problems (Group V): 2, 14, 13, (11).
High on Graphing (Groups III and IV): 10, 4, 17.

Relating Testing and Assignment Content

All questions used by teachers in quizzes, classroom tests, and exams were categorized. From teacher-supplied marking schemes and weighting systems, the relative (percentage) weight of every question in the calculation of the final grade was determined. These relative weights were summed over the 126 topics (not reported), the 14 Content Categories (Appendix E), and the six Content Groups (Table 6). A comparison of corresponding percentages in Tables 5 and 6 reveals a general trend toward less testing than assignment of Group I (basic skill questions), slightly more testing than assignment of Groups II and III (proofs and differentiation graphing), and considerably more testing than assignment of Groups IV and V (integration graphing and situational problems). At least some of this pattern — how much our data do not reveal — can be traced to differences between the size of questions for basic skills and the size of questions for the other groups of content. The emphasis on Group I content in assignments may also reflect a belief in something like "practice makes perfect" rather than a belief that algorithmic routines are of overwhelming importance. Moreover, the greater emphasis on Group II content in testing than in assignment may mean that proofs are considered important, but are dealt with by class talk and demonstration rather than assignment.

****

Insert Tables 6 and 7 about here

****

Table 7 was produced by subtracting the cell entries of Table 5 from the corresponding cell entries of Table 6. Positive numbers in Table 7 indicate greater testing emphasis than assignment emphasis. A comparison of the numbers in the column of Table 7 indicates the extent to which the testing-assignment difference varied over teachers.

The most striking feature of Table 7 is the fact that all the numbers in the column for content Group I are negative. Teachers uniformly tested basic skills less than they emphasized the skills in assignments, although the difference for some teachers (2 and 10, in particular) was less than for others (Teacher 5, in particular). For the other content groups, testing emphasis was typically greater than assignment emphasis, with only a few large, positive differences to be found.

Effect of exemptions. Apart from the main data set, we also examined the effects of exemptions from final exams and dropping of test results in the calculation of final grades. The motivation for detailed investigation of these situations is that the content basis for comparing students will not remain constant if students are assigned grades based on different tests and/or exams. Table 16 contains data on the impact of exempting students with a high enough term mark (typically 65%) from the final exam. Four teachers followed such a policy, as mandated by either the school or the board. The results in Table 8 reveal that three of the four teachers were able to compare students on roughly the same content basis, whether or not they took the final exam. That is, the final exam appears to reflect the content that was tested for the term mark. The exception is Teacher 5, whose final exam appears to have reflected Content Groups I and VI more and Content Group IV less than the term tests. Note that under the exemption policy, Group IV for Teacher 5 is weighted more than twice as heavily than it is for any other teacher in the sample.

****

Insert Table 8 about here

****

Effect of optional test dropping. Table 8 also contains data on Teacher 16, who followed a policy of allowing students to drop one poor test result from the calculation of final marks. Teacher 16 divided the semester into five terms. Each term contained up to six short quizzes and one test. The tests for terms 3 and 5 were considered to be the midterm and final exam respectively, but students were allowed to drop the test for one of terms 1, 2, or 4, along with the associated quizzes from the calculation of the final mark. The data in Table 16 show that dropping the test for Term 1 or Term 2 has about the same impact, a shift in emphasis from Group I to V in comparison with the situation in which all tests count. The dropping of the test for Term 4, however, shows an opposite trend, that is, a shift away from Group V.

Examination of Appendix E for Teacher 16 reveals differences in content categories, depending on which term test is dropped. The dropping of Term 1 marks almost eliminates the evaluation of skill in limits (Content Category 1) from the final grade. The dropping of Term 4 marks reduces to about one-third the effect of assessments of integration skills (Category 4); and the dropping of Term 4 marks eliminates entirely from the final grade the evaluation of Category 8 content (finding areas between curves) and Category 12 content (max/min problems). If the basis for student grades is to remain approximately constant over students, to facilitate comparisons of relative achievement, then the policy of allowing the results of some tests to be dropped from the calculation of final grades must be accompanied by a policy of carefully balancing the content of each term test. But balancing the content of term tests seems an unreasonable requirement.

A *micro-analysis*. Reporting of the analysis of test and assignment content in six content groups ignores much of the fine detail present in the topic-level data. To show this, data from a subset of five teachers on Content Category 10, motion problems, were analyzed. The teachers who were chosen had each assigned a moderate number of questions for Category 10. The topics for Content Category 10, along with the corresponding data for this sub-sample of teachers, are given in Table 9.

Table 9 is a summary of these data in two different ways. Table 9B gives the total numbers of questions assigned and marks awarded. For the entire category, assignment totals are fairly constant, as would be expected given how these teachers were selected, but testing weights for the category vary from more than 10% of the entire calculus course to only 1.5%. Examination of each of the seven topics covered in Table 9B reveals considerable variation in teacher assignments. All teachers assigned questions for Topics 2220, 2230 and 2240, but two or three or four of these teachers assigned no questions for the other four topics. Clearly, our consideration of only the sum of topic entries in Category 10 has masked some differences in content coverage and perhaps also in the difficulty of the content being covered.

****

Insert Table 9 about here

****

Table 9C gives percentages of the raw totals for each teacher, as reported in Table 9B. Table 9C reveals very large discrepancies between content as assigned and as tested. The data must be interpreted, however, in the knowledge that these seven topics form at least a rough hierarchical sequence, and that failure to test a lower level topic may be justified. In this context, the following observations were made:

— Teacher 8 taught but did not test Topics 2205 and 2210; Teachers 12 and 13 tested but did not assign questions for Topic 2205; Teacher 15 neither assigned nor tested both topics, and Teacher 12 neither assigned nor tested Topic 2210.

— Teachers 3 and 15 neither tested nor taught Topics 3172 and 3173, while Teacher 12 dealt with these topics in assignment but not in testing. Teachers 8 and 13 neither tested nor taught Topic 3172, and Teacher 11 dealt with it in assignment but not in testing.

This short analysis reveals that the reporting in large content categories and groups has the effect of masking differences that appear at a more detailed level of analysis. From some perspectives (e.g., that of a first year physics professor faced with a mixture of students who may or may not have been exposed to trigonometric motion), this is a serious limitation.

Grading Practice

Work in this part of the report was directed by four research questions about the grade assignment practices of the 17 teachers:

1. Did the process of assigning grades differ from teacher to teacher?
2. Did the grades assigned vary?
3. Were the differences in processes related to the differences in grades assigned?
4. What questions are suggested for systematic research on grade assignment?

The Process of Assigning Grades

The following characteristics of process were considered: the grading policies established by the board and school administrations, the criteria used to compute student grades and the weights of each, the computational methods used to derive grades, the time intervals at which data for grading were collected throughout the course, and the features of the term tests and final examinations.

Grading Policies. The teachers were following policies set at various administrative levels, from the provincial Ministry of Education to the mathematics departments within the 17 schools. Given that Ministry policy was constant, variations in the practices described reflect variations in school- or board-level policies, and do not necessarily reflect variations in the teachers' personal philosophies.

By provincial policy, every student must write a minimum of one examination. The format and the weight of that exam, however, is determined by policy of the board, school, or mathematics department. In six cases, teachers reported that the policy was set by the board. For four of these cases, the boards allowed schools to exempt from the final examination those students who had maintained a 65% average throughout the course. All students in these schools wrote a mandatory mid-term exam to satisfy provincial policy, and a small percentage of the students in each class — those not exempted — wrote the final exam. One board required that students take a compulsory board-wide final exam worth either 35% or 65% of the student's final grade, the percentage depending on the student's exam mark relative to his or her term mark. The higher of the exam mark or the term mark was weighted 65%. In the sixth case, the board stipulated that the weight of the final exam could range from 20% to 40% of the final grade.

Three teachers reported that the exam policy originated at the school level. Two schools specified that the maximum weight of the final exam should be 35%. Another school specified that the weight of the final exam should be 40%.

The remaining eight teachers attributed the exam policy to their department. In seven cases, the stipulated weight of the exam was either 40% (5 teachers), 35% (2 teachers) or 25%. In the case of the eighth teacher, the mathematics department followed an experimental grading policy in which the course was divided into five month-long segments. Students received a grade for each segment based on the results of a one-hour exam, quiz scores and a mark for participation. Exams in segments 3 and 5 were considered the mid-term and final exams, respectively. The lowest grade obtained in the remaining three segments, 1,2 or 4, was eliminated from the calculation of the final grade.

One other policy should be noted as having originated at the mathematics department level. In each case, all calculus classes in a school wrote the same final examination. The setting of examinations was the responsibility of the mathematics department. The department head usually assigned one teacher the task of preparing an exam, with the exam content being approved by the other calculus teachers. Provided there was more than one calculus teacher, different teachers prepared the mid-term and the final exams.

The provincial policy on grading in pre-university mathematics courses will change over the next year as the new OAC (Ontario Academic Course) mathematics curriculum (Ontario Ministry of Education, 1985) is adopted across the province. The new policy will require that the student grade be composed of a minimum of 40% for formal examination results. This will alter some of the grading practices reported here.

Grading Criteria. The seventeen teachers used 22 different grading systems. Five teachers employed two different systems in their classes. Four of these teachers instituted exemptions policies, so the grading system for each student depended on whether the student had been exempted from the final examination. In the table of results for grading criteria (Table 10), the grading systems for non-exempted students are designated by the subscript 'n' and those for exempted students by the subscript 'x'. The fifth teacher (#16) used an experimental grading policy in which the lowest of three term grades (representing 20% of the course) was dropped from the calculation of the final grade. Thus, '16a' refers to the situation in which the dropped term grade was from the first half of the course, and '16b' to the situation in which it was dropped from the second half.

****

Insert Table 10 about here

****

Table 10 lists the grading criteria and the percentage weight of each criterion for each of the 22 cases. Grading criteria consisted almost entirely of two forms of testing: term tests, usually administered at the end of a unit of work, and examinations, administered either at the mid-point or at the end of the course. For 7 of the 17 teachers, tests and exams represented 100% of the students' grade. For the other 10 teachers, additional criteria included quizzes (6 teachers, weight ranging from 2% to 20%), assignments (6 teachers, weight ranging from 3% to 6%), and a subjective mark for participation (4 teachers, weight ranging from 5% to 20%). (Participation was the teacher's assessment of the student's attitude and, in one case, also depended on attendance.) Six of the teachers gave no mid-term exam, while the four with exemption policies had no final exam for the exempted students. The remaining grading systems included both mid-term and final exams.

There was considerable variation over teachers in the number of criteria considered. To obtain a sense of the extent of this variation, consider Teachers 4 and 11. Teacher 4 used seven criteria, composed of six term tests and a final exam, while Teacher 11 used 15 criteria, consisting of 13 term tests, an assignment, and a final exam. The final exams of each teacher were worth 40% of the grade, and term work 60%. Due to the different number of term tests, Teacher 4's students were earning 10% of their final grades when they wrote such a test, while Teacher 11's students were earning less than 6% of their final grades per term test.

For five teachers, the grading criteria differed from student to student in the same class. In the classes of the four teachers with exemptions policies (1, 5, 6, 7), the term tests written in the second half of the course were weighted more heavily (as a percentage of the final grade) for students exempted from the final exam than for students not exempted. For students in the class of Teacher 16, the elimination of a first-half test from the calculation of the final mark meant that the weight of the first-half test score was 50%. The elimination of a test from the second-half meant the weight of the first-half test scores was 75% of the final grade. (We use the designations first-half and second-half rather than term 1 and term 2 to avoid confusion. All our data were collected in semestered schools in the February to June semester. The labels semester and term are usually synonymous. The break between first-half and second-half is about mid-April, depending on the school.)

Teachers assigned different relative weights to these grading criteria. Term test weights ranged from 30% to 80% of the grade. The total number of term tests ranged from 3 to 13. For the 22 grading systems, the weight of first-half tests averaged 40%, with a range from 23% to 75%. Course examinations (midterms and finals combined) were assigned weights from 20% to 60%. On average, the overall weight of the examination component was higher when two exams rather than one were used (46% compared to 38%).

Computational Methods. The teachers used combinations of several methods to aggregate scores on different criteria in computing final grades. Three different methods of aggregation were used:

1. Differential or equal weighting of tests. The tests were weighted according to the number of marks in each (simple summation of scores on all tests) (Teacher 2, 3, 8 [first half only], 9, 11, 13) or they were re-weighted to counteract differences in the number of marks, the re-weighting being done so as to make the weights of the tests equal (weighted summation of scores) (Teacher 1, 4, 5, 6, 7, 8 [second half only], 9, 12, 14, 15, 16) or so as to reflect the teacher's perception of the relative importance of the topics covered by a test (Teacher 10, 17).
2. Differential versus common grading criteria for all students within the class. Grades were based either on the same criteria for all students in a class, or on a configuration of criteria that varied from student to student and that depended on the student's performance. Examples of varied criteria are these: the lowest test score(s) for each student was dropped from the computation of a grade (Teacher 11, 16); students were allowed to rewrite a different version of the test on which they had achieved their lowest score (Teacher 13); a make-up test on several topics was administered at the end of the first half, but was used only if the score on the test would improve the student's grade (Teacher 8); and students were exempted from writing the final exam if their term mark was a minimum of 65% (Teachers 1, 5, 6, 7).
3. Weighting the two halves of the semester. The course is divided into two halves, each of about nine weeks duration. Five teachers weighted each half equally (1n, 1x, 3, 5n, 5x, 6x, 7x). For Teachers 6 and 7, the marks for the two halves for students in the exempted groups were equally weighted, while the marks for the two halves for students in the non-exempted groups were weighted 4:6 and 3:7 respectively. For Teacher 16, the weight of first-half marks was .75 for students who dropped a

monthly grade from the second half, otherwise the marks for the two halves were equally weighted. For the other eleven teachers, the average weight of first half marks was .31.

Management of Testing Time. In Table 11 total testing time is reported for each teacher. The time spent on testing activities (including time for midterm and final exams) ranged from 9 to more than 17 hours. Over the length of the course, an average of 8.8 hours (range 3.0 hours to 15.2 hours) was spent in writing an average of 8 class tests (range, 3 tests to 13 tests). The length of class tests varied from 25 to 75 minutes. With the exception of Teacher 16, quizzes contributed very little to the calculation of grades. Eleven of the 17 teachers (including the four with exemptions) administered a midterm exam, and all administered a final exam (but not to all students in the case of teachers with exemptions).

****

Insert Tables 11 and 12 and Figure 1 about here

****

Table 12 indicates the length of each teaching segment that preceded a test. Nine of the 17 teachers (Teachers 1, 2, 8, 11, 12, 13, 14, 15, 17) seemed to have more regular cycles of teaching and testing than the other teachers. In general, the other eight teachers tested fairly consistently throughout the first half of the course, but their patterns became erratic in the second half. For example, for four teachers (3, 6, 9, 16), the lengths of the first teaching segments prior to testing in the second half was from 19-24 hours, about twice the average length of first-half teaching segments prior to testing. For Teacher 4, the last teaching segment in the second half (prior to the final examination) was 20 hours long, compared to an average of 11 hours for previous segments. The teachers who followed more regular teach-testing cycles also gave a greater number of tests on average (10 compared to 7). Differences among selected teachers are portrayed graphically in Figure 1.

Item Frequency. The number of test items on each term test is listed in Table 13. The total number of items for which the 17 teachers awarded term marks averaged 104 per teacher, with a range from 53 (3 term tests for Teacher 16) to 180 (10 term tests for Teacher 1).

****

Insert Table 13 about here

****

Topic Coverage. Another factor to consider in the grading process is the degree to which the grading criteria are cumulative measures of achievement. One index of cumulative testing is the number of different topics that are covered in the same test paper. For this analysis, the taxonomy of calculus course content described in Table 3 was used. It should be noted that teachers are not expected to have taught all 14 content categories. Two of the 14 (no. 13 and no. 14) form Content Group 6, a unit that is taught in one of the three Grade 13 mathematics courses. Which course these topics is placed in is usually determined by the mathematics department in a school. Content Category 9 is the third optional category not covered in the 1972 Grade 13 Calculus Guideline.

Table 14 is a summary of the number of content categories covered by the tests administered by each teacher. (See also Appendices F and G.) The number of categories on each term test ranged from one to eight. For each teacher, the average number of categories per test ranged from 2.0 to 4.7 (Teachers 5 & 16). All but two teachers (4, 17) gave at least one test covering only one category.

****

Insert Table 14 about here

****

For five teachers, the final exam included content categories that had not been covered in class tests. In three cases, the weights of these topics on the final exam were substantial: 24%, 28% and 84% (Teachers 5, 4, 16). One possible reason for Teachers 4 and 5 is that the final examinations used by these teachers were set by other calculus teachers. When more than one teacher in a school teaches a course and the classes are to write a common final exam, a certain amount of prescience is required to select items for which all students will have had an equal opportunity to learn. A class that proceeds more slowly than expected is likely to be disadvantaged by the exam.

Teacher 16's low test-exam topic overlap is easily explained. This teacher did not discriminate between tests and exams; the 3 term tests and 2 exams each represented one month's content coverage, and were not cumulative.

The Assigned Grades

The following discussion of grades is focused on four considerations:

1. the differences between midterm and final course grades,
2. the effect of differential grading systems on the grades assigned in the four classes with exemptions policies,
3. the weight of the final examination in the grades, and
4. the correlations between term tests and final examinations.

Data for this part of the paper were collected from the teachers mark-sheets, on which were recorded all test scores and other data used in the calculation of grades, as well as the midterm and final grades.

Midterm and Final Course Grades. The Calculus course is divided into two halves, and percentage grades are assigned at the end of the first-half (the midterm grade) and at the end of the course (the final grade). The average midterm and final grades for each class, and the difference between the two is given in Table 15. The distinction between midterm and final grades is important in Ontario, particularly for Grade 13 courses offered in the Spring semester (February through June), as was the course considered in this study. Early in April, Ontario universities begin their first-year admissions process, and most of the students who take Calculus are applicants. For the courses a student is taking in April, the school submits interim grades (normally the same as midterm grades) to the university applications centre, and students receive a conditional acceptance or rejection from the university based on these grades. A concern expressed by several teachers in the study is that students become less motivated to work once the interim grades have been submitted.

****

Insert Table 15 about here

****

Averaged over all 17 teachers, the final grade of 67 was 6 percentage points lower than the average midterm grade of 73. For every teacher, the mean final grade was either lower than or equal to the midterm grade. The range of mean final grades was 54 to 75, that of midterm grades, 61 to 79. Although there is a high correlation, on average, between the two sets of grades for a teacher, the mean difference between the midterm and final grades varied from 0% to 15%. While slackening motivation is a possible explanation for lower final exam marks, it is also possible that final exams either contain harder questions, have broader content coverage, or are graded less generously than tests earlier in the term.

Final Examination Weights. Whatever the reasons, there are additional consequences of final exams having lower marks, especially in those classes where only some students are required to write final exams. In Table 16, the mean final exam scores and term marks of each class are listed with their standard deviations. There are two aspects of note in these data. First, final exams tend to be harder than term marks. For the students in classes with no exemption policy, the difference was about 12 marks; for the students who were in classes with exemption policies and who had to write the final exam, the difference was 19 marks. Second, final exams tend to discriminate more than term marks. For the classes with no exemption policy, the standard deviation of final marks was about 50% larger than that of term marks. For the students writing finals under an exemption policy, the standard deviations of final exam marks were twice as large as those for term marks. Thus, the much harder final exams had considerably more impact on the discriminations made among students by the final grades than was intended by the teachers when they decided the nominal weighting of final exams and term marks.

****

Insert Table 16 about here

****

Test-Examination Correlations. Correlations between test scores and the final exam scores were examined for twelve of the seventeen teachers. Teachers 1, 5, 6, 7, and 16 were omitted because they represent exceptional cases, with exemptions or experimental grading policies). Table 17 contains the correlations for each class test and the average correlation, over class tests, for each teacher. Average correlations ranged from .61 to .93, with an overall average of .74. Half of the teachers had fairly consistent correlations on all tests. The low correlations for six teachers (2, 10, 12, 13, 15, 17) may, however, indicate that certain tests were unusual in terms of their content:

-- For Teacher 2, the correlation between Test 5 and the exam is .50. The content categories on the test were Curve sketching and Optimization Problems. On the exam, the first of these content categories represented 9% of the available marks, but the second content category was not examined.

-- For Teacher 10, Test 3 correlated .55 with the exam. The only content category covered on the test was Optimization Problems, which represented 6% of the final exam.

-- For Teacher 15, Test 4 correlated .37 with the exam. The content category covered in the test was Motion Problems, but this material was not covered on the final exam.

In all three of these cases, the tests with low correlations covered types of situational problems (motion, related rates and optimization). Situational problems might be perceived as being generally more difficult than other types of questions.

****

Insert Table 17 about here

****

Relating Testing and Grading Processes to Grading Data. Across the seventeen teachers in the study, little evidence has been found to suggest that differences in process are related to differences in grades. However, this analysis is based on only a small number of teachers. In a different study, less intensive but involving more teachers, the following questions might be considered:

1. Does the total number of term tests administered by a teacher reflect differences in the design and content coverage of those tests?
2. Do teachers who administer a large number of test items emphasize different course topics (e.g., more skill-level questions) than teachers who administer a small number of items?
3. Do specific topics tend to appear on single-topic tests?
4. When the number of content categories tested on one test is large, how are the topics related?
5. Do teachers use particular designs of tests for all their tests, or is the design of a test related to the topic being covered?
6. Do teachers whose midterm and final grades are about the same assign a higher weight to tests from the second half of the course?
7. Does the difference between intended and actual exam weights reflect a difference between the content coverage of the tests and the exam?
8. For teachers whose tests are not consistently correlated with the final exam, how do tests with low correlations differ from tests with high correlations?

We turn now from an examination of testing practice to investigation of marking standards.

Responses to the Common Marking Task

Recently, Milton, Pollio and Eison (1986) argued that to understand grades and the grading process, it is necessary to comprehend the context in which grades are assigned. This context is multifaceted. It consists of the social and historical setting of the era in which the grade was assigned, the traditions of the educational institution involved, the traditions of the academic discipline involved, the characteristics of the students being graded, and the grading philosophy and policies of the instructor. In the study reported here, the social and historical context is that of Ontario in the mid-1980s, and the grading of mathematics for senior high school, university-bound, academic students.

The results of this project, as described thus far, provide only a loose basis for comparing the marking standards of the 17 teachers. All were offering what was ostensibly the same calculus course, but, as demonstrated in the earlier sections of the report, the courses differed in non-trivial ways. The academic traditions of the 17 secondary schools at which the teachers worked varied to a greater or lesser extent. The characteristics of the students being taught were different from school to school. The approaches to instruction and the content of that instruction varied over teachers. And student evaluation practices varied. But we find no basis in the evidence presented thus far for answering such questions as these:

— To what extent do the 17 teachers agree in their assessments of the importance of given examination questions for eliciting evidence of student achievement?
— To what extent do the teachers agree on the quality of student responses to given examination questions?
— To what extent do the teachers agree as to which kinds of examination performances represent failing achievement and which represent passing achievement?

These were the sorts of questions addressed in this phase of our study.

The purpose, then, was to investigate the extent and nature of the variation among 17 teachers of calculus in the standards they applied in marking a set of examination papers. Variation was expected, so from the outset the study was intended to describe the nature of that variation and relate it if possible to whatever else was known about the teachers — their philosophies of teaching and marking, their understanding of the curriculum to be taught, their attitudes toward examinations, their conceptions of what can and should be examined, their impressions of the particular examination used in this study, and their impressions of the students whose papers were being marked.

Background

The literature on educational measurement contains a number of reports documenting variation in the marking standards of teachers. The classic investigations in the United States by Starch and Elliott (1912, 1913a, 1913b) are often cited. So too is the work in England of Hartog and Rhodes (1935, 1936). Studies of factors associated with variation in the marking of essays has been reported over the years in the Journal of Educational Measurement (Chase, 1968, 1979, 1986; Daly & Dickson-Markman, 1982; Hales & Tokar, 1975; Hughes, Keeling & Tuck, 1980b; Hughes & Tuck, 1984; Marshall & Powers, 1969) and in other journals (Hughes, Keeling & Tuck, 1980a, 1983; Rafoth & Rubin, 1984). It is clear from these reports that different readers are likely to assign different marks to the same essay examination paper, even when the same marking guide is applied by all readers. It is also well-known that when asked to re-mark a paper after a period of time, a reader will often assign it a different score. The issue here is not to re-confirm the existence of variation in marking standards. The issue instead is to continue work in the tradition of Hartog and Rhodes (1935, p. 10), who called for "... careful and systematic experiment [so] that methods of examination can be devised not liable to the distressing uncertainties of the present system." We do not expect that the well-known faults of essay grading, including exercises and problems that are commonly used to assess mathematics achievement, can be overcome easily, if at all. Our modest hope is that information about variation in standards will provide a foundation for the development of procedures and materials that teachers can use to bring their differing standards more nearly into line.

Data and Analysis

Several kinds of data specific to the issue of variation in marking standards were collected. A final calculus examination, administered in June 1987 to students in a school not otherwise involved in the study, yielded, after sorting and culling, a set of 20 papers that spanned a range of quality. The 17 teachers who participated in this study were given the 20 papers, and each was asked to prepare a marking guide and then mark the papers against it. In addition, the teachers were asked to provide written comments, should they care to make any, about the examination and performance of the students. The resulting data were analyzed in a variety of ways, both quantitative and qualitative. A description of each type of analysis is provided when it is first encountered in the remainder of this report.

Results

What follows has been divided into five main sections, dealing, respectively, with the examination, the marking guides, the total marks assigned students, the individual questions in the examination, and teacher impressions and comments not covered in the other sections of this report. The last section of the paper contains a summary and discussion of the problem of increasing the consistency of teacher grading standards.

The Examination

The exam contained 11 questions, reported in Appendix H. Those numbered 1, 4, 7, 9, 10, and 11 did not include separately designated sub-questions. The remaining questions — 2, 3, 5, and 6 — consisted of three or more separately lettered sub-questions. Number 2 had 13 parts, 3 had nine, and 5, 6 and 8 had three parts each. In total, the examination consisted of 37 questions and sub-questions.

The content of the each question can be described briefly as follows:

1. A quadratic function is defined and the examinee is asked to find the point on the function where the tangent has a specified slope.
2. The examinee is asked to obtain the derivative of y with respect to x for each of 13 different functions of x. Of these functions, six involve the logarithmic or exponential function, four a polynomial function, and three a trigonometric function.
3. Nine parts which test integration: three of these sub-questions are trigonometric functions, four are logarithmic or exponential functions, and two are polynomial functions.
4. Tests the ability to integrate using the method of parts.
5. Three limit sub-questions; each limit involves polynomials.
6. A motion problem, with sub-questions involving a) acceleration, b) velocity, and c) the position of the particle in motion after a specified amount of time has elapsed.
7. The examinee is directed to find the area enclosed between two trigonometric functions of the same variable over a specified range of the variable.
8. A cubic function is defined and the examinee is asked to do three things: a) find the coordinates of all maximum and minimum points of the function, b) find the coordinates of all points of inflection, and (c) sketch the function.
9. An application problem: the examinee is to find the rate at which the distance between two moving objects is increasing or decreasing, given information about the direction and rate of motion of the two objects.
10. Asks for a proof that the formula given in the question for the volume of a sphere can be obtained as the volume of revolution of a circle of specified radius.
11. The examinee is asked to find the radius and height of a cylinder, such

that the cylinder will have a given volume and an unspecified but minimum surface area.

The questions and sub-questions of the examination were classified according to the scheme developed in this study for categorizing homework and test questions. The results, with the numbers of sub-questions to those questions with parts given in parentheses, are as follows:

	Category	Exam Questions
I.	Basic skills (differentiation, integration)	2 (13), 3 (9), 4, 5 (3)
II.	Proofs	10
III.	Investigating functions using differentiation	1, 8 (3)
IV.	Investigating functions using integration	7
V.	Situational problems / applications	6 (3), 9, 11

Using number of questions and sub-questions as the basis for judgement, we can see that the exam is strongly weighted toward the testing of basic skills. (The foregoing classification is not wholly satisfactory in that the categories are not mutually exclusive. For example, Question 10 involves finding a volume of revolution, and can be placed in Category IV as well as Category II.)

The Marking Guides

Initial analysis. The marking guides, which teachers prepared without recourse to any information other than that conveyed by the examination paper itself and the written responses of 20 students, indicate the maximum number of marks to be awarded for responses to each question. A complete listing of the allocations of marks by the 17 teachers to all questions and sub-questions is contained in Table 18.

****

Insert Table 18 about here

****

Scrutiny of the first line of Table 18 reveals considerable variation among teachers in the total number of marks allocated for perfect performance of the examination. This fact is highlighted in a stem-and-leaf display of the total numbers of marks allocated by the 17 teachers:

Stem	Leaf
14	4
13	0
12	013356
11	69
10	35569
9	6
8	
7	8

As can be determined from this display, the smallest of the maximum marks was 78, the largest 144, and the median 116.

The differences among teachers in the total number of marks reflects the fact that they also differed in their allocations of marks to individual questions and sub-questions. For example, as Table 18 will confirm, wholly satisfactory performance of Question 7 was rewarded with as many as 12 marks by two teachers, and as few as 4 marks by one teacher. True, an acceptable answer to sub-question (h) of Question 2 was allocated two marks by every teacher, but this consistency was not found for any other question or sub-question. The inconsistency found for the third part of Question 2 is perhaps more typical; an acceptable answer to this question was assigned six marks by three teachers, five marks by two teachers, four marks by seven teachers and three marks by five teachers.

What accounts for these differences among the teachers? For the most part, they seem to stem from differences in the number of steps or stages to an answer that are awarded marks. For example, the first question on the exam asked students to "[f]ind the point on the curve y = 4x2 + 2x - 3 at which there is a tangent with slope -14." Without exception, the teachers gave as the model solution one involving the following steps:

(i) Obtain the derivative of the function.

(ii) Set the derivative equal to -14.

(iii) Solve the resulting equation for x.

(iv) Substitute the value of x into the original equation and evaluate y.

(v) State that the obtained values of x and y define the point required in the question.

The number of marks awarded for answers agreeing with the model answer ranged from three to five. The teachers who gave five marks awarded one mark for successful completion of each of the five steps noted in the model solution. Those who awarded only three marks gave one mark for correctly reaching steps (iii), (iv) and (v); or (i), (ii) and (v); or (i), (iii) and (iv); or (i), (ii) and (iv). Various combinations were also noted in the marking patterns of the teachers who gave four marks for a model answer to Question 1. And one teacher imposed a one mark penalty for failing to begin an answer by restating the function for the curve. Appendix I contains sample marking schemes for questions 1, 9, and 10, as produced by Teachers 1 and 16. (Appendix I also contains sample student responses to be discussed later.)

A further source of variation in the marks awarded for answers was noted for Question 7. Here examinees were to "[f]ind the area enclosed between y = 2sin(x) and y = sin(3x) for [the range] 0 [less than or equal to] x [le] pi." The model solution required students to set the two functions equal and solve for the points of intersection over the range of x specified in the question. Having determined these points, the model solution then involved integrating over three separate portions of the range of x, two where y = sin(3x) lies above y = 2sin(x) and one where y = 2sin(x) is higher. Six of the teachers and all but four of the students assumed that one of the functions was higher than the other over the whole of the specified range of x. The resulting wrong answer involved far less mathematics and many fewer steps than the correct answer. The six teachers who marked Question 7 against the incorrect answer key allocated fewer marks in their marking guides — on average only six — than the teachers who marked against the correct answer — nine marks on average. Appendix I contains marking schemes for question 7 as produced by Teacher 13, who used an incorrect solution, and for Teacher 14, who used a correct solution.

Another difference already alluded to was in the use of bonus marks and deductions. Several marking guides indicated bonus marks for good form and for stating the answer in a complete English sentence. Several others indicated deductions for failing to include the constant of integration in answers to integration questions or for failing to specify units in answers to questions involving measured quantities. These bonuses and deductions, when used, were either one mark or one-half mark.

A final difference of note is idiosyncratic to the marking scheme in which separate marks were awarded for method and accuracy. This is the marking scheme of Teacher 10, whose total number of marks was largest. It is possible that other teachers also distinguished in some marks between method and accuracy, but none did so explicitly on the marking guide.

Despite the obvious disparities found among the teachers' marking schemes, an issue that merits consideration is the extent to which teachers were in agreement as to the order of importance of the examination questions and sub-questions. This issue is addressed in the next two sub-sections of the paper. In the analyses reported in these sub-sections, Question 8 was treated as unitary, with only the total mark used in the analyses. It was necessary to proceed in this way because six teachers did not assign separate marks to the parts of Question 8.

Principal components analysis. Consider Table 19. Each element in this table is the coefficient of correlation for a pair of teachers between the maximum marks allocated by each teacher to the 35 questions and sub-questions of the examination. All the intercorrelations are substantial, ranging from 0.76 to 0.95, with a median of 0.89. Clearly, there is a high degree of agreement among the teachers in their perceptions of the relative importance of the questions and sub-questions in the examination. (This impression of agreement is a little misleading in that a minority of questions and sub-questions were assigned large numbers of marks relative to the majority of questions and sub-questions, hence the minority of questions and sub-questions had a larger influence on the correlation coefficients than the majority.)

****

Insert Table 19 about here

****

The coefficients of correlation in Table 19 were further explored using principal component analysis. As would be expected, given the consistently high magnitude of the correlation coefficients in the table, its first principal component accounted for 89 percent of the variance among teachers in their allocations of marks to questions and sub-questions. The second principal component accounted for only 3 percent of the variance. The teacher coefficients on the first principal component ranged from a low of 0.90 to a high of 0.98, where a coefficient of 1.00 is the maximum possible. The square of one of these coefficients represents the proportion of variance in the corresponding teacher's allocation of marks that can be accounted for by the first principal component. These proportions are very large, ranging from 0.81 to 0.96.

We should consider whether the second principal component, despite the small amount of variance associated with it, provides meaningful distinctions among the teachers, and adds substantially to the account of the correlation table given by one component. It does not. Only six teachers (1, 7, 9, 11, 15, 16) had second component coefficients greater than 0.20 in absolute value, and the largest of these was 0.30. (Later it is shown that Teacher 1 was the hardest marker and Teacher 16 was the easiest, but this polar opposition seems not to account for the fact that they both had negative coefficients on the second component, whereas the other teachers listed had positive coefficients on this component.)

The results of the principal component analysis were used to provide a sense of the relative weight given the examination questions in the various

Content Groups on Table 7. This was done by calculating the scores of the questions and sub-questions on the first principal component. The sum of the scores for the questions in each group of questions represented on the examination is as follows:

	Content Group	Exam Questions	Sum
I.	Basic skills	2, 3, 4, 5	924
II.	Proofs	10	72
III.	Investigating functions using differentiation	1, 8	198
IV.	Investigating functions using integration	7	115
V.	Situational problems / applications	6, 9, 11	295

It is clear from these results that the teachers' marking schemes weight the examination heavily toward basic skills, with the questions in this content group counting more toward the final examination mark than the questions in all other groups combined.

Generalizability Study. A two-way analysis-of-variance of the teacher marking schemes yielded additional evidence of agreement among the teachers in their perceptions of the relative importance of each question and sub-question on the examination. Making the dubious assumptions that the 17 teachers were selected at random from an infinite population of teachers and that the 35 questions and sub-questions were chosen at random from an infinite population of questions, we obtain the following estimates of the components of variance for these data: Questions 4.13, Teachers 0.17, Interaction 0.77. Clearly, the main source of variation in the numbers comprising the body of Table 18 is the questions factor. The interaction factor stands a distant second as a source of variation. [An interaction here implies that the relative order of importance of questions and sub-questions varies in a way not perfectly explained by the relative positions of question means (over teachers) together with the relative positions of teacher means (over questions).]

Very little of the variation in the numbers in Table 18 is explained by differences among teachers in the average number of marks allocat- per question.

Translated into reliability terms, these results suggest the following: In their allocations of marks to questions and sub-questions, the teachers distinguished very reliably among questions (generalizability coefficient of 0.84), but those same allocations of marks did not yield reliable distinctions among the teachers (generalizability coefficient of 0.18).

The overwhelming impression left by the results of the quantitative analyses of the teacher marking guides is that the teachers possessed very similar views of the relative importance of the questions and sub-questions of the examination.

Qualitative Analyses. Although the teachers were agreed on the relative importance of the questions in the examination, they did not unanimously agree that the exam gave good coverage of the calculus course described in the Guideline (Ontario Ministry of Education, 1972). Several teachers (9, 10, 13, 14 16) objected to the strong emphasis in the exam on integration. Three teachers (4, 9, 16) noted the lack of coverage of polar coordinates and complex variables. Two teachers pointed to the coverage in the exam of trigonometric functions, with one (# 1) feeling it was inadequate and another (# 9) thinking it was overemphasized. It was observed by Teachers 9 and 16 that volumes of revolution, trigonometric limits and differentials were given short shrift. And three teachers (10, 12, and 16) objected to the preponderance of skill-type questions, and the lack of questions involving problem-solving. Note that volume of revolution, polar coordinates, and complex numbers are optional topics.

In a draft document entitled *A Handbook for the Examination Component of Evaluation in the OAC — Calculus* (Ontario Ministry of Education, 1987), attention is paid to the number of marks awarded for arithmetic and algebraic simplification in answers to OAC calculus examination questions. An analysis was made of the marking guides in an attempt to assess the extent of differences among them in the proportions of marks awarded for arithmetic, algebraic simplification, and other skills and knowledge (from earlier grades) compared to the calculus skills and knowledge to be acquired in the course. (This analysis was possible for 13 of the 17 guides; four guides indicated only total numbers of marks per question or sub-question, and did not indicate how the marks were to be distributed over the parts of model answers.) The 13 percentages of marks for calculus as opposed to other kinds of mathematical knowledge and skill ranged from 60 to 76, with a median percentage of 66. Thus, there was some variation from teacher to teacher in the extent to which knowledge and skills peripheral or prerequisite to the calculus were rewarded.

Total Student Marks[2]

Principal factor analysis. The marks assigned each paper by each teacher were analyzed using principal factor analysis. As in the principal component analysis of the marking guides, this analysis proceeded by treating each teacher as a variable, thus yielding a table of intercorrelations among teachers — Table 20. The numbers in Table 20 are uniformly high. (They range from 0.81 to 0.97, with first quartile, median and third quartile equal to 0.90, 0.92 and 0.94 respectively.) Obviously, there is close agreement among the teachers in the relative orders into which they placed the 20 sets of responses to the examination.

****

Insert Table 20 about here

****

A one-factor model fits the correlations in Table 20 very well. The single factor accounts for 92 percent of the variance in the teachers' marks. The coefficients of the teachers on the factor range from a low of 0.90 to a high of 0.99. If the squares of the coefficients are interpreted as estimates of the reliability of marking — these numbers are communalities, hence estimate a lower bound to reliability — the estimated reliability coefficients range from 0.81 to 0.97.

---

[2] There were 12 addition errors, by five different teachers, in the 340 final marks submitted. Nine of the errors were less than 5% of the total percentage marks for the papers, but the other three were 9%, 10%, and 17% respectively.

A two-factor model was also fit to the correlations in Table 20, but this model adds little to the account of the correlation table that is provided by the one-factor model. This is not surprising in that the second, un-rotated factor accounts for less than two percent of variance in the teachers' marks. In addition, the second factor, even when rotated, gave no hint of meaningful differences among the teachers.

Generalizability analysis. An analysis of variance of the total marks assigned by each of the 17 teachers to the 20 student papers produced the following estimates of components of variance: for students 122.7, for teachers 116.1, and for interaction 15.1. The important feature of these results is that the component of variance for interaction is small relative to the size of the component of variance for students. This reflects the high degree of agreement among teachers in their rank ordering of the examination papers. It also reflects the relatively high reliability attached, on average over teachers, to the examination score assigned a paper by a teacher; the coefficient of generalizability for decisions about the relative merits of the student papers is 0.89.

The results reported thus far support the strong conclusion that the 17 teachers were in close agreement in their rankings of the 20 examination papers. But there is more to the story than this, because the component of variance for teachers in the generalizability analysis was relatively large, indicating very different levels in the overall marking of the teachers.

Percentage marks and absolute standards. Grading achievement in calculus and other subjects involves more than rank- ordering a group of students. Determinations of fail and pass and honours are usually required. How well, then, did the teachers agree as to which papers represented failing performance, which represented passing performance, and, of the passes, which represented honours. To address this question, the total mark a teacher assigned a paper was converted into a percentage of the total mark given in the marking guide. These percentages are reported in Table 21.

****

Insert Table 21 about here

****

The bordering columns and rows of Table 21 provide evidence of inconsistency in standards. The papers of Students 10 and 20 were assigned percentage marks of 80 or more (honours) by 13 and 12 teachers, respectively, but the other teachers assigned these papers nothing more than passing marks, ranging from 67 to 79. The papers of Students 4 and 16 were assigned percentage marks below 50 (failing) by seven and nine teachers, respectively; the other teachers assigned marks ranging from 50 to 61. Teachers 1, 9 and 14 assigned no paper a mark in the honours range, and Teacher 1 assigned failing marks to seven papers. On the other hand, Teachers 2, 4, 5, 8, 11, 12 and 16 assigned no paper a failing mark, and Teacher 16 assigned percentage marks of 80 or more to 10 papers. Variation in standards is apparent, despite the fact that the teachers ranked the papers for quality in very much the same way.

The median polish procedure of exploratory data analysis (Tukey, 1977) was applied to Table 21. The teacher effects, which are here defined as the differences between the median marks of individual teacher and the median mark over all teachers — the common effect — confirm that Teacher 1 was the hardest marker, awarding an average of 15 percentage points less per paper than Teachers 10 and 13, who centered the scale with marking effects of zero. Other hard markers, relatively speaking, were Teachers 3 (-5.5 points), 9 (-6), 14 (- 8) and 15 (-6.5). The easiest marking teacher was number 16 (+8 points), followed by Teachers 5 (+6) and 12 (+6). From the perspective of the examination papers, the median polish procedure gave a scale centered on the papers of Students 11 and 17, with computed effects of zero. As would be expected from the numbers in Table 21, the papers of Students 4 and 16 were generally low with effects of -20 and -19, respectively. The papers of Students 6, 7, 9, 12 and 19 were also comparatively low, with effects of -10, - 9, -11, -7 and -9, respectively. The students whose papers were associated with positive effects were 10 (+16) and 20

(+12), in line with the impression conveyed by Table 21, followed by 5 (+9), 15 (+6) and 18 (+9).

The large residuals of the median polish procedure identify percentage marks not closely reproduced by adding the corresponding student and teacher effects to the common (overall) effect (here, 70). Although over 50 percent of the residuals of the median polish procedure were less than 2 in absolute value, they did range from -8 to 9 (standard deviation of about 3). Of 340 residuals, 12 were outliers (as defined in exploratory data analysis — Tukey, 1977). Teacher 1 was associated with a disproportionately large number of the extreme residuals — five, of which three were negative. Teachers 9 and 11 were each associated with two large residuals, and Teachers 3, 4, and 17 were each associated with one. Only the paper of Student 1 (effect of 2) was associated with as many as two outlying (negative) residuals.

As evidence that, on average, an additive model gives a reasonably good account of the percentage marks data, despite the 12 large residuals, we note that the ratio of the inter-quartile range of the distribution of residuals to that of the original marks is only about one-fifth (3.5 to 17).

Teacher comments. The teachers offered comments, several of which are of interest in view of the results presented in the previous sub-section. For example, Teacher 1, the stiffest of the markers, directed comments at student performance: the solutions were poorly developed, diagrams were missing, and the responses lacked clear, concise statements. These might be described as errors of form in the student responses. (Teacher 1 penalized poor form. Although the marking guide of this teacher indicated five marks for the first question on the examination, no student was awarded more than three marks for his/her answer. The apparent reason for this was the failure by all 20 students to include all the steps listed in the teacher's model answer. Thus, for example, no mark was awarded for finding the y-coordinate of the answer if the determination of this coordinate had not been made explicit, even when the student's answer did contain the correct coordinate.) Teacher 14, another of the hard markers, also noted the errors of form, but so did Teachers 6 and 7, both of which had marker effects near the central (zero) point. The added fact that Teachers 3, 9 and 15, the others with relatively large, negative marking effects, did not mention form of answer as a problem, suggests that this factor may not fully explain the large, negative marker effects.

In fact, no comment appears to distinguish the hard from the easy markers. The easiest marker, Teacher 16, described the marking exercise as boring. This teacher also described the exam questions as being all of the skill and recall type, and as not requiring higher level thinking skills. It is not apparent that adopting this point of view should cause one to be an easy marker, although Teacher 12, another easy marker (effect of 6), also commented on lack of problem solving questions on the exam. So too, however, did Teacher 10 (marker effect of 0). Perhaps more in line with what might be expected, given his/her large negative marker effect, was Teacher 14's registration of disappointment in the students' poor problem solving abilities. Teacher 4 (with a marker effect of 3) thought the exam was too easy. Teachers 6 and 9 (effects of -1 and -6) said the exam was of uneven difficulty, with questions being either very easy or very difficult. Teacher 6 thought the exam was too long, but then so too did Teacher 5 (effect of 6). That the latter also felt the exam was "too tricky by half" may go some way to justifying his/her easy marking. The foregoing results should be contrasted with those for Teachers 3, 7, and 8, who expressed the view that the exam achieved a nice balance between straight-forward and challenging questions. The marker effects for these teachers were -5.5, -2.5 and 2, respectively.

There is a factor that does appear to distinguish hard from easy marking teachers to an extent. Ten of the teachers followed one textbook (published by Gage) and six others followed another textbook (published by Holt). (One teacher used a set of notes, and followed no published book.) Seven of the Gage teachers were associated in the median polish analysis with positive marking effects (median marking effect over the 10 teachers: 2.5). Five of the six Holt teachers were associated with negative marking effects and the effect associated with the sixth teacher was zero (median over the six teachers: -4.5). Further clues to differences among the teachers in their marking standards were sought in an analysis of the marks assigned individual examination items.

Marks on Individual Items

A study was made of the marks assigned responses to individual questions. In one analysis, an item-total correlation was computed for each exam question or sub-question for each teacher between the marks that that teacher had given on the 20 papers for answers to a question and the total marks he/she had assigned the papers. In another analysis, components of variance were estimated for the factors of a three-way fully-crossed design, the factors consisting of teachers, students and items.

We begin this discussion, however, with a presentation of exemplary data. The top half of Table 22 reports the marks awarded to three students (4, 10, and 17 — a weak, top, and average student respectively) by two teachers (#1, the hardest marker, and #16, the easiest marker) for answers to three questions (#1, discussed earlier in detail, #9, and #10). The bottom half of Table 22 contains marks awarded to several students by Teachers 13 and 14, the first working from an incorrect model answer and the second from a correct model answer. Actual student responses and teacher marking schemes are reported in Appendix I. For question 9, note two instances where two students received the same mark from one teacher but different marks from the other. For question 10, students 4 and 17 received almost the same mark from Teacher 1, but very different marks from Teacher 16. Finally, for question 7, there are many large discrepancies between the marks awarded by the two teachers.

****

Insert Table 22 about here

****

Discriminating items. A summary of the item-total correlations obtained for each item is contained in Table 23. Included in the table are several statistics pertaining to the distribution of 17 coefficients (one per teacher) for each question: the median, the minimum, and the maximum. Several sub-questions, specifically 2a, 2e, 2m, 5a and 5c, were poor discriminators in that as many as 16 of the teachers had assigned all papers the same mark for answers to these sub-questions. Other sub-questions — 2d, 2i, 3g, 3h, and 6b — were poor discriminators despite the fact that not all were assigned the same mark. It is interesting to note that all of the most poorly discriminating sub-questions except 6b fell in the category of skill questions. Among the better discriminating items were 2b, 2f, 2j, 3c, 3d, 3f, 3i, 4, and 10. Note again that all but Question 10 were skill-type questions. Also note that with the exception of 2b, 2f and 2g, these questions involved integration. This suggests that the skill of integration had not been assimilated as uniformly well by the students as the skill of differentiation.

****

Insert Table 23 about here

****

The teacher comments are interesting to consider in connection with these results. Several of the sub-questions that did not discriminate well were described by at least six teachers as being too simple (2a, 2m, 5a and 5c) or too difficult. It was further suggested that the difficulty of some questions is due to the excessive sophistication required or the need to employ a trick in deriving the correct answer (2i, 3g). Of the better discriminating of the differentiation questions, 2b was described as tricky but OK, 2f as a good test of simplification (although at least one teacher objected to the large amount of simplification required), and 2j as a tough but reasonable question. Among the better discriminating of the integration questions, 3f was described by 12 teachers as difficult to the point of being too hard and unfair. Several also stated they did not think 3f covered knowledge included in the Ministry Guideline

Generalizability analysis. In an attempt to isolate sources of variance in the marks assigned by the 17 teachers to the 20 papers, a three-way generalizability analysis was performed, with the 35 questions and sub-questions as the third facet of the analysis. The components of variance for this and two other analyses are reported in Table 24. The striking feature of the results for the analysis involving all 35 questions and sub-questions is the fact that the large variance components were associated with main and interaction effects for questions. By comparison, the components of variance for the teacher main effect, the student main effect and the student by teacher interaction were small.

****

Insert Table 24 about here

****

The results for all 35 questions enable us to calculate two coefficients of generalizability, that for relative and that for absolute decisions about students (Brennan, 1983). Not surprisingly, the coefficient for relative decisions, 0.72, is considerably larger than that for absolute decisions, 0.31, in line with previous observations. (These coefficients assume a student's mark is based on performance of 35 questions marked by one teacher.) The challenge is to devise a marking scheme that would deliver substantially greater reliability of absolute decisions.

The large components of variance in Table 24 are those involving questions. The effect of these components can be attenuated by increasing the number of questions on the examination and by constructing the exam according to a stratified domain sampling plan.

Consideration can also be given to reducing the teacher components of variance. There are two ways in which these teacher sources of variation might be reduced. The simplest and most direct way is to have more than one teacher mark each examination paper, awarding the paper an average of the marks assigned by the different teachers. (Were each exam paper marked by three teachers, the generalizability coefficient for absolute decisions would be predicted to increase for the 35 question exam from 0.31 to 0.49, neither a startling nor by itself a satisfactory increase.) The other way to attack variability among teachers is to try to enhance consistency in marking. Where this is likely to have most effect is suggested by the results of the generalizability analyses conducted of two subsets of the 35 questions. These analyses were of (i) the questions classified as testing basic calculus skills (26 of 35) and (ii) all the other questions. It can be seen from the results presented in Table 24 that the variance components for the skills questions are considerably smaller than those for the other questions.

The results of the latter analysis were followed up by considering the marks assigned by four teachers — the hardest and easiest markers (1 and 16), the teacher at the centre (#13), and a teacher who was as hard a marker, relatively speaking, as Teacher 16 was an easy marker (#14) — to the three students already discussed (see Table 22) with respect to individual items. It can be seen in Table 25 that the teachers differed relatively little in the percentages of the total marks each allocated for the 26 basic skill questions that were given for the responses of a student, although as would be expected, there was a tendency for Teacher 16 to be somewhat more generous than the others. Against this standard, the corresponding results for the non-skills questions are dramatically different. There are dramatic differences among the teachers in the percentages of the marks allocated to the non-skills questions that were awarded a particular student's responses. For example, Student 4 was awarded about half the marks allocated by Teachers 13 and 17 to the other-than-basic skills questions, but Teachers 1 and 14 assigned only one-fourth the marks they had allocated for performance of the same questions. It appears that the main source of the difference among these teachers in grading standards lies in their marking of the exam questions that test other-than-basic differentiation and integration skills.

****

Insert Table 25 about here

****

Teacher comments. Teacher comments about the exam fell into three categories -- positive, negative and neutral. The 10 teachers who used the Gage text had varied views of the exam, as has been noted, with four of the 10 teachers being classified as positive, four negative and two neutral. The six teachers who used the Holt text were either negative or neutral. This difference in attitude may have stemmed in large part from the view the teachers had of the match of the exam to the Ministry Guideline for the calculus course. Most of the Holt teachers commented somewhere in their marking guides that the exam did not reflect the Guideline very well, but none of the Gage teachers mentioned the match of the exam to the Guideline.

Summary

This was an intensive study of a small number of teachers. It was designed to explore, in depth, some un-examined aspects of classroom assessment, and on the basis of this exploration, formulate questions to be addressed in future studies. The main results of each of the four parts of the study can be summarized as follows:

Use of Classroom Time

The analysis of use of classroom time showed: (a) total class time required for the calculus credit ranged from 96 to 114 hours (median 105), (b) time spent on testing, excluding formal examinations, ranged from 9 to 17.2 hours (median 12.5), (c) percentage time spent on teacher-centred direct instruction ranged from 17% to 52% (median 26%), (d) percentage time devoted to seat-work ranged from 8% to 46% (median 26%), and (e) percentage time devoted to review ranged from 4% to 14% (median 11%). Clearly, the teachers differed substantially in their organization for teaching. Moreover, those with more class time available expended a smaller percentage of time on direct instruction, and allocated greater percentage to homework and practice.

Content Coverage

The purpose of the content coverage analysis was to examine the degree to which marks assigned by different teachers can be said to reflect the same content basis. Contrary to what might be expected for a group of highly experienced and qualified teachers of mathematics, an area often considered to be assessed with considerable consistency as compared to many others (e.g., English literature), the results show substantial differences in content emphasis. Apart from differences in grading practice, we can conclude that the calculus backgrounds of students from the classrooms of our teacher volunteers would differ substantially, depending in part on classroom attended. To summarize:

- Teachers varied widely in the number of questions assigned as homework — from under 500 to more than 1600. A large part of this variation was accounted for by differences in the number of questions on basic skills.
- Teachers varied considerably in the extent to which they emphasized different topics in their assignment of questions. For example, the number of questions on basics ranged from 40% to 75% of the total number of assigned questions.
- The emphasis on basic skills was less in the tests than the assignments, whereas the emphasis on other content groups was greater in the tests than the assignments.

One approach to homogenizing student experiences in calculus is to encourage dialogue among teachers. Which teachers ought to be involved in this dialogue is, however, problematic. If the question is comparability of the backgrounds of students enrolled in first-year university classes, then discussion at the school or board level, while perhaps relatively easy to arrange, would be unsatisfactory. The alternative, discussion at the provincial level, besides being difficult in practice, might require a change in teachers' views of centralized curricular control.

Our results show differences within classes between the content emphasis of assignments and the content emphasis of tests. Some of these within-teacher differences between assignment and testing emphases are probably intentional as when a teacher decides to test only at the top of a small hierarchy of skills or knowledge, ignoring the prerequisite skills and knowledge that had been included in assignments. Conversely, a teacher may teach a difficult concept and choose not to test it because most students failed to grasp it. Whether or not discrepancies between teaching and testing constitute a problem to be corrected is a matter not addressed in the present study. All we have done here is provide evidence to make mathematics educators aware that such discrepancies as these exist.

This study leaves unanswered several questions pertaining to content coverage. We have been hampered substantially by our necessary reliance on teacher-provided logs, and by the obviously limiting assumption that assigned questions provide an accurate representation of course content.

The Grading Process

From the analyses of the grading practices of the 17 teachers, it was learned that examinations and term tests were the two main determinants of student grades. Seven of the 17 teachers used exam and test results exclusively as grading criteria; the other 10 teachers augmented these results, to a minor extent, with other information. However, the main determinants of grades were operationalized in different ways in different classes and, in five cases, in different ways for different students within the same class.

All students in all classes wrote a minimum of one examination, as required by provincial policy. However, the experience of taking an examination varied for students in different classes. For the four classes with an exemption policy, the majority of students took their only exam on material learned in the first half of the semester. For six other classes, the only exam was a final exam based on the entire semester's work. In the remaining seven classes, both a mid-term exam and a final exam were required. The time that students from different classes spent in an examination situation ranged from 2 to 4.5 hours. The final examination mark was weighted from 15% to 40% of the student's final grade and, when a mid-term exam was administered, the resulting mark was weighted 9% to 30% of the final grade. In some cases, the calculus content topics that were tested during the semester were not emphasized to the same extent on the final exam. This might be attributed to the fact that, while setting of term tests is an individual teacher's responsibility, the final examination is a mathematics department's responsibility and is usually set by only one of the calculus teachers, not necessarily the teacher who participated in the study.

Term testing was found to vary in the following respects: (i) number of tests (ranging from 3 to 13), (ii) number of items comprising the tests (from 53 to 180), (iii) amount of classroom time used for test taking (from 3 to 15.2 hours), and (iv) schedule of tests (sporadically or regularly). Term tests were weighted from 30% to 80% of the final grade.

Ten teachers considered factors in addition to exam and test scores in determining grades. Other measures of cognitive skills included quizzes, administered by six teachers and weighted from 2% to 20%, and take-home assignments or class presentations, given by six teachers and weighted from 3% to 6%. Four teachers included a participation mark, based on such non-cognitive measures as attendance, effort or attitude, and weighted from 5% to 20%.

The variations in grading process described in this report cannot be evaluated positively or negatively from the evidence obtained. Thus, none of the grading processes described herein can be recommended for adoption at a regional or provincial level. Still it is evident from this study that students taking Grade 13 Calculus in the Spring 1986 semester from the 17 teachers in this study did not demonstrate their achievement in calculus through a common process of assessment and grading. It is reasonable to question whether or not it would be beneficial for students to have experienced similar grading processes, and to have been judged according to similar standards on similar criteria of achievement.

Following from this limited study of the grading process, several questions may be asked about the appropriateness of setting more uniform grading standards for this course. First, as regards examination practices:

1. What are the implications for final grades of the final examination being weighted 15% versus being weighted 40% of the final grade? Should every student in this course write a final examination carrying the same weight toward the final grade?
2. How cumulative should a final examination be? Should students be examined on their achievement of only that content taught during the

second half of the course or should they be examined on the content of the entire course?

3. Should the content emphasis on a final examination mirror the content emphasis on term tests?
4. Should every student write a midterm examination carrying the same weight toward the final grade? Does the experience of writing a midterm examination help students develop more realistic expectations of a final examination?

Second, on testing practices:

5. Should every student write the same number of term tests? What are the implications of one class writing three term tests, and another writing thirteen?
6. What are the implications of students being required to answer 50 items as opposed to three or four times as many on term tests?
7. Should teachers be testing at regular intervals throughout the course? What are the effects of irregular testing on student learning and motivation?
8. Should term testing be cumulative throughout the course? What are the implications for student learning and retention of testing on discrete topics as opposed to multiple topics? of testing students on recent as opposed to not-so-recent learning?

Finally, on other grading criteria:

9. Should teachers of an academic, pre-university course be crediting attendance or participation in class?
10. What other grading criteria should be considered, criteria that could foster types of mathematical skill not necessarily assessed through a traditional testing situation?

There is a need to study the implications of various grading processes for impact on student learning and in terms of the feasibility and efficiency.

The following questions illustrate the types of inquiries needed on the effect of grading processes on learning and retention:

11. Are different learning styles reinforced by different grading methods?
12. Does a concentration on single topics or single units in tests foster reliance on short-term memory and limit the development of understanding?
13. Does cumulative testing throughout a course help students synthesize course material and retain knowledge?
14. Does a regular schedule of testing help students develop better study habits?

On the implications of grading for expectations of future success:

15. How do high school grading practices affect students' expectations of grading at the college or university level?
16. Are students from certain high schools more successful in college and university because characteristics of the grading systems in these schools closely resemble the grading systems of the tertiary institutions?

The Marking Process

The main results of the empirical study of the marking process were as follows:

1. There was substantial agreement among the 17 teachers as to the relative importance of the examination questions. The examination did not cover all facets of the calculus course, however, so it cannot be inferred that the same degree of agreement would have been realized had the exam sampled the course domain more extensively.
2. There was substantial agreement among the teachers as to the relative quality of the 20 student papers that were marked. Another way of describing this result is to suggest that in judging the quality of student responses to calculus examinations, these teachers used much the same ordinal or interval scale of measurement.
3. There was substantial disagreement among the teachers as to the absolute quality of the 20 student papers. Continuing the metaphor of the previous

point, these teachers were in marked disagreement as to where the zero point on the scale for judging calculus achievement should be placed.

4. There was the suggestion in the results that the marking standards of teachers varied to a limited extent as a function of the textbook being used in the course.

These results pose a challenge for the Ministry of Education in a jurisdiction where there is no external mechanism — no common, province-wide examination — for aligning standards for pass-fail judgements of calculus achievement. This challenge has not been lost on critics of education in Ontario, and it has not been ignored by the Ontario Ministry. *A Handbook for the Examination Component of Evaluation in the OAC — Calculus* (Ontario Ministry of Education, 1987), the draft handbook referred to earlier, was developed for the purpose of fostering a greater degree of consistency in calculus examinations across the province. The following matters were discussed in the handbook:

1. The practice of exempting selected students from final examinations, a practice that fosters inconsistency in the bases on which the achievement of different students in the same class is assessed.
2. The undesirable effect of varying the weight of the final examination, either for different students within the same class or from one class (school, local educational authority) to another.
3. The importance of assessing problem solving in addition to fundamental knowledge and skills.
4. The need for consistency in marking schemes — (i) in the award of marks for arithmetic and algebraic simplification instead of, or in addition to, the award of marks for calculus, (ii) in the award of partial marks for correct procedures, and (iii) in the imposition of penalties for errors of form and for the omission of work that can be done mentally.
5. The desirability of following the same examination rules governing the use of calculators and the provision of measurement formulas.
6. The potential value of a common examination for all the schools in a local education authority.

In-service training sessions addressing these matters were held for department heads and superintendents responsible for calculus courses across the province. Attention was focussed in these sessions on the examination practices and exams and marking schemes that had been found in an earlier Ministry survey to be in use in the province.

The Ministry draft handbook and in-service program address several problems found to exist in the present series of studies — (i) the practice of granting exemptions from final examinations and the variation in value of final examinations, (ii) the emphasis on basic skills to the virtual exclusion in teaching and testing of problem solving, and (iii) the wide differences in amount (by time and number of questions) of testing and other assessment activities. But the results of the present study suggest that consistency in assessment will be increased only when other important steps are taken as well.

1. There is need to establish consistency in the way marks are awarded for performance of examination questions testing other-than-basic calculus skills. Procedures for marking, and the development of training modules by which the procedures may be practiced, need to be devised for accomplishing this.
2. There is need to increase consistency in the way teachers assign marks for displays on examinations of other-than-calculus knowledge and skills.
3. There is need for more than one teacher, ideally for several teachers, independently to mark student exam papers. Student marks can then be based on the average of the different marks, ideally after large differences have been discussed and resolved.
4. There is need to adopt stratified schemes for sampling course content when preparing exams, and there is need to ensure that exams are sufficiently long or sufficiently numerous that the impact on student grades of differences among questions in their relative importance is reduced.
5. Following up on an observation made in this study, a systematic investigation needs to be made of the relationship that exists, if any, between textbooks used and teacher marking standards.

References

Alexander, D. W. (1987). *Evaluation practices: Calculus. Interim report.* Toronto, Ontario: Ministry of Education.

Anderson, J. O. (1987). *Teacher practices in and attitudes towards student assessment.* Paper presented at the Annual Meeting of the Canadian Educational Researchers' Association, McMaster University, Hamilton, Ontario. June, 1987.

Bloom, B. S. (Ed.) (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain.* London: Longmans.

Brennan, R. L. (1983). *Elements of generalizability theory.* Iowa City, Iowa: ACT Publications.

Burstein, L. (1983). A word about this issue. *Journal of Educational Measurement, 20*, 99-101.

Chase, C. I. (1968). The impact of some obvious variables on essay test scores. *Journal of Educational Measurement, 3*, 315-318.

Chase, C. I. (1979). Impact of achievement expectations and handwriting quality on scoring essay tests. *Journal of Educational Measurement, 16*, 39-42.

Chase, C. I. (1986). Essay test scoring: Interaction of relevant variables. *Journal of Educational Measurement, 23*, 33-41.

Daly, J. A., & Dickson-Markman, F. (1982). Contrast effects in evaluating essays. *Journal of Educational Measurement, 19*, 309-316.

Fleming, M., & Chambers, B. (1983). Teacher-made tests: Windows on the classroom. *New Directions for Testing and Measurement, 19*, 29-38.

Graham, N. (1984). *Designing and marking English examinations - A resource booklet for Scarborough English teachers.* Toronto, Ontario: Scarborough Board of Education.

Green, K. E., & Stager, S. F. (1986). *Effects of training, grade level, and subject taught on the types of tests and test items used by teachers.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco.

Haertel, E. (1986). *Choosing and using classroom tests: Teachers' perspectives on assessment.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Hales, L. W., & Tokar, E. (1975). The effect of the quality of preceding responses on the grades assigned to subsequent responses to an essay question. *Journal of Educational Measurement, 17*, 131-135.

Hartog, P., & Rhodes, E. C. (1935). *An examination of examinations.* London: Macmillan, 1935.

Hartog, P., & Rhodes, E. C. (1936). *The marks of examiners.* London: Macmillan.

Herman, J. L., & Dorr-Bremme, D. W. (1983). Uses of testing in the schools: A national profile. *New Directions for Testing and Measurement, 19*, 7-17.

Hughes, D. C., Keeling, B., & Tuck, B. F. (1980a). Essay marking and the context problem. *Educational Research, 22*, 147-148.

Hughes, D. C., Keeling, B., & Tuck, B. F. (1980b). The influence of context position and scoring method on essay scoring. *Journal of Educational Measurement, 17*, 131-135.

Hughes, D. C., Keeling, B., & Tuck, B. F. (1983). The effects of instructions to scorers intended to reduce context effects in essay scoring. *Educational and Psychological Measurement, 43*, 1047-1050.

Hughes, D. C., & Tuck, B. F. (1984). The use of model essays to reduce context effects in essay scoring. *Journal of Educational Measurement, 21*, 277-281.

King, A., Clements, J., Crawford, D., Good, H., Lapierre, A., Macintosh, D., MacNab, S., Revill, A., Russell, H., Stroud, T., Warren, W., & Wilson, I. (1976). *Continuity and diversity of courses. The secondary/post-secondary interface project III: Nature of Programs.* Toronto, Ontario: Ministry of Education and Ministry of Colleges and Universities.

Lazar-Morris, C., Polin, L., Moy, R., & Burry, J. (1980, August). *A review of the literature on test use.* Los Angeles: Center for the Study of Evaluation. CSE Report No. 144.

Marshall, J. C., & Powers, J. C. (1969). Writing neatness, composition errors, and essay grades. *Journal of Educational Measurement, 6*, 97-101.

Marso, R. N., & Pigge, F. L. (1988). *An analysis of teacher-made tests: Testing practices, cognitive demands, and item construction errors.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans.

Milton, O., Pollio, H. R., & Eison, J. A. (1986). *Making sense of college grades.* San Francisco: Jossey-Bass.

Ontario Ministry of Education. (1972). *Mathematics: Senior division.* Toronto: Government Printer.

Ontario Ministry of Education. (1985). *Mathematics: Intermediate and senior divisions. Part three.* Toronto: Government Printer.

Ontario Ministry of Education. (1987). *A handbook for the examination component of evaluation in the OAC — Calculus (DRAFT).* Toronto.

Rafoth, B. A., & Rubin, D. L. (1984). The impact of content and mechanics on judgments of writing quality. *Written Communication, 1*, 446-458.

Rudner, L. M. (Ed.). (1984). *Testing in our schools: Proceedings of the NIE invitational conference on test use.* Washington, D.C.: National Institute of Education.

Starch, D., & Elliott, E. C. (1912). Reliability of the grading of high-school work in English. *School Review, 20*, 442-457.

Starch, D., & Elliott, E. C. (1913a). Reliability of grading work in mathematics. *School Review, 21*, 254-259.

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

Stiggins, R. J. (1988). *The nature and quality of teacher-developed classroom assessments.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans.

Stiggins, R. J., & Bridgeford, N. J. (1985). The ecology of classroom assessment. *Journal of Educational Measurement, 22*, 271-286.

Stiggins, R. J., Conklin, N. F., & Bridgeford, N. J. (1986). Classroom assessment: A key to effective education. *Educational Measurement: Issues and Practice*, *5* (2), 5-17.

Terwilliger, J. S. (1987). *Classroom evaluation practices of secondary teachers in England and Minnesota.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, Washington, D.C.

Tukey, J. (1977). *Exploratory data analysis.* Reading, Mass.: Addison-Wesley.

Wahlstrom, M. W., & Danley, R. R. (1976). *Assessment of student achievement.* Toronto, Ontario: Ministry of Education.

Wilson, R. J. (pending). *The processes of evaluating student achievement: Two case studies.* Queen's University, Kingston, Ontario.

TABLES

Table 1: Descriptive Characteristics of Classes and Schools

Teacher	Class Size	School Size (nearest 50)	Teaching Hours (incl. exams)	Number of Tests
1	17	1400	108	10
2	30	950	103	8
3	13	300	107	7
4	15	1200	97	6
5	18	1100	96	6
6	28	1200	105	7
7	22	900	106	6
8	19	1900	105	10
9	16	1900	103	7
10	25	1000	111	10
11	24	2050	105	13
12	31	1800	103	10
13	27	1500	110	9
14	21	1250	105	9
15	13	1300	104	8
16	26	1400	103	3
17	27	950	114	9

Table 2: Percentage of Time Spent on Classroom Activities

Teacher	Administration	Direct Teaching	Review Work	Review Test	Homework	Practice Seatwork	Practice Review	Assessment[1]
1	0.6	51.8	8.7	3.0	10.3	8.7	0.9	16.2
2	0.1	30.1	6.2	6.6	22.3	23.5	0.4	10.9
4	0.0	38.9	5.0	3.7	11.9	31.3	0.0	9.3
6	0.5	25.3	6.0	6.5	27.4	15.9	8.5	9.9
7	0.1	24.0	6.3	1.9	26.1	28.4	5.5	7.8
10	1.0	33.9	7.3	2.4	11.3	30.2	0.5	13.3
11	0.3	18.6	9.1	4.7	43.0	8.0	0.4	16.0
13	0.8	27.9	1.0	2.8	15.2	34.5	8.2	9.7
14	5.3	29.5	7.8	3.1	17.3	22.6	3.9	10.6
15	0.3	20.8	4.7	4.8	29.3	21.8	3.9	14.6
16	1.9	24.2	1.3	4.0	11.5	46.0	1.0	10.2
17	1.7	17.1	7.9	0.6	18.4	36.6	8.7	9.2
Mean	1.1	27.9	5.9	3.7	20.[illegible]	25.6	3.5	11.5

[1]Assessment includes all tests and examinations

The results in Table 2 are based on an analysis of teachers' logs. The percentages given here of time for assessment activities vary as much a 3% from the times reported in Tables I and II, which were based on additional information.

Table 3: System of Groups and Categories of Calculus Topics

Derived from a Taxonomy of 126 Topics (See Appendix B.)

I Skill-level Questions:

1 = Limits
2 = Sequences & Series
3 = Differentiation
4 = Integration

II Proofs:

5 = First Principles and Other Proofs

III Questions Involving Graphing -- Differentiation:

6 = Slope/Equation of a Tangent
7 = Curve Sketching

IV Questions Involving Graphing -- Integration:

8 = Finding Area Between Curves
9 = Finding Volume of Revolution[1]

V Situational Problems:

10 = Motion Problems
11 = Related Rates
12 = Maximum/Minimum (Optimization)

VI Optional Units:

13 = Complex Numbers
14 = Polar Coordinates

[1]Category 9 is optional; unlike Categories 13 and 14, which may be optional in the sense that may be taught in other senior mathematics courses, the topics in this category can not be taught in another Grade 13/OAC mathematics course.

Table 4: Numbers of Questions Assigned by Content Groups and Content Categories within Groups[1]

Teacher	Content Groups/Categories: I				II	III		IV		V			VI		Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
1	95	60	386	258	1	133	61	47	39	38	120	55	213	116	1622
2	58	41	251	115	2	37	56	39	5	107	87	15	5	18	836
3	110	60	597	207	1	112	87	67	14	78	70	33	4	58	1498
4	16	21	153	50	1	55	40	36	7	29	39	9	0	0	456
5	64	30	526	117	0	62	13	71	14	25	16	19	11	35	1003
6	24	0	180	92	0	29	18	34	1	35	8	39	0	0	460
7	76	64	307	157	13	42	44	32	0	90	60	32	0	0	917
8	153	180	529	186	24	66	86	11	0	72	56	25	0	0	1388
9	62	41	407	178	3	71	128	113	36	37	64	89	133	33	1395
10	0	0	346	199	3	147	177	112	88	145	45	51	0	28	1341
11	57	28	273	193	20	69	92	48	0	85	65	37	0	0	967
12	138	0	450	174	3	51	59	59	0	73	48	28	0	0	1083
13	72	46	212	118	0	69	65	59	45	75	61	70	78	0	970
14	84	9	307	155	1	76	120	49	0	137	34	65	0	0	1037
15	104	0	492	287	7	51	90	119	39	66	51	37	0	0	1343
16	66	20	164	175	18	35	78	40	4	50	85	17	59	39	850
17	27	44	441	176	4	137	98	78	24	128	23	93	1	0	1274
Total	1206	644	6021	2837	101	1242	1312	1014	316	1270	932	714	504	327	

[1] Includes both homework and seat-work

Table 5: Percentage of Assignment Questions by Content Group

Teacher	Content Group					
	I	II	III	IV	V	VI
1	49.3	0.1	12.0	5.3	13.1	20.3
2	55.6	0.2	11.1	5.3	25.0	2.8
3	65.0	0.1	13.3	5.4	12.1	4.1
4	52.6	0.2	20.8	9.4	16.9	0.0
5	73.5	0.0	7.5	8.5	6.0	4.6
6	64.3	0.0	10.2	7.6	17.8	0.0
7	65.9	1.4	9.4	3.5	19.8	0.0
8	75.5	1.7	11.0	0.8	11.0	0.0
9	49.3	0.2	14.3	10.7	13.6	11.9
10	40.6	0.2	24.2	14.9	18.0	2.1
11	57.0	2.1	16.6	5.0	19.3	0.0
12	70.4	0.3	10.2	5.4	13.8	0.0
13	46.2	0.0	13.8	10.7	21.2	8.0
14	53.5	0.1	18.9	4.7	22.8	0.0
15	65.7	0.5	10.5	11.8	11.5	0.0
16	50.0	2.1	13.3	5.2	17.9	11.5
17	54.0	0.3	18.4	8.0	19.2	0.1
Median	55.6	0.2	13.3	5.4	17.8	0.1

Table 6: Percentage Test Marks by Content Group

Teacher	Content Group					
	I	II	III	IV	V	VI
1	34.9	2.5	14.0	12.3	19.6	16.6
2	49.6	3.0	12.9	3.3	20.0	11.1
3	50.8	6.7	16.5	4.1	15.3	6.5
4	33.0	6.8	18.4	11.6	30.2	0.0
5	43.0	2.2	12.7	15.7	19.9	6.5
6	49.6	2.7	13.1	13.4	21.1	0.0
7	46.5	1.6	18.3	8.0	25.6	0.0
8	58.1	1.3	17.7	0.0	23.0	0.0
9	37.1	2.7	13.6	11.4	21.5	13.7
10	32.8	2.3	23.2	12.4	27.1	2.3
11	34.6	5.5	27.2	8.2	24.2	0.0
12	48.1	7.5	14.8	8.1	21.5	0.0
13	26.1	3.5	18.9	14.6	27.5	9.4
14	32.8	5.4	23.9	6.7	31.3	0.0
15	48.1	5.7	14.5	13.5	18.3	0.0
16	37.6	3.5	14.9	6.2	25.3	12.6
17	34.6	10.2	15.4	14.4	25.3	0.0
Median	37.6	3.5	15.9	11.4	23.0	0.0

Table 7: Testing Emphasis minus Assignment Emphasis[1]

	Content Group					
Teacher	I	II	III	IV	V	VI
1	-14.4	2.4	2.0	7.0	6.5	-3.7
2	-6.0	2.8	1.8	-2.0	-5.0	8.3
3	-14.2	6.6	3.2	-1.3	3.2	2.4
4	-19.6	6.6	-2.4	2.2	13.3	0.0
5	-30.5	2.2	5.2	7.2	13.9	1.9
6	-14.7	2.7	2.9	5.8	3.3	0.0
7	-19.4	0.2	8.9	4.5	5.8	0.0
8	-17.4	-0.4	6.7	-0.8	12.0	0.0
9	-12.2	2.5	-0.7	0.7	7.9	1.8
10	-7.8	2.1	-1.0	-2.5	9.1	0.2
11	-22.2	3.4	10.6	3.2	4.9	0.0
12	-22.3	7.2	4.6	2.7	7.7	0.0
13	-20.1	3.5	5.1	3.9	6.3	1.4
14	-20.7	5.3	5.0	2.0	8.5	0.0
15	-17.6	5.2	4.0	1.7	6.8	0.0
16	-12.4	1.4	1.6	1.0	7.4	1.1
17	-19.4	9.9	-3.0	6.4	6.1	-0.1
Median	17.6	2.8	3.2	2.2	6.8	0.0

Table 8: Effects of Exemptions and Differential Test Weighting[1]

Teacher	Content Group I	II	III	IV	V	VI
<u>Data on Exemption Policy</u>						
1_n	34.9	2.5	14.0	12.3	19.6	16.6
1_x	35.2	1.6	14.0	11.2	23.2	14.8
5_n	43.0	2.2	12.7	15.7	19.9	6.5
5_x	38.1	2.1	11.9	29.6	18.5	0.0
6_n	49.6	2.7	13.1	13.4	21.1	0.0
6_x	47.2	3.7	12.9	13.2	23.0	0.0
7_n	46.5	1.6	18.3	8.0	25.6	0.0
7_x	44.3	2.7	16.9	8.7	27.4	0.0
<u>Data on Teacher 16, Three Different Test Dropping Options</u>						
All inc.	37.6	3.5	14.9	6.2	25.3	12.6
Drop T1	31.2	1.9	14.5	7.6	29.5	15.4
Drop T2	29.9	2.4	12.8	7.7	31.5	15.7
Drop T4	39.0	4.6	17.9	3.2	18.8	16.5

[1]values are precentages of marks assigned to each group

Table 9: Motion problems (Content Category Ten) as Treated in Assignments and Tests by a Six Teachers

A. Topics in Category 10

2205. Inst. vs. avg. rate of change, related problems
2210. Find velocity from speed, time, using graphs and tables
2220. Find velocity as a derivative
2230. Acceleration: simple graphical and computational practice
2240. Acceleration and closely related problems
3172. Motion problems involving quadratic relations
3173. Motion problems involving trigonometric functions.

B. Numbers of Questions Assigned (Q) and Marks Awarded (M)

Teacher	3		8		11		12		13		15	
Topic	Q	M	Q	M	Q		Q	M	Q	M	Q	M
2205	23	0.6	13	0.0	14	0.0	0	2.6	0	2.3	0	0.0
2210	4	0.0	6	0.0	10	0.3	0	0.0	7	0.0	0	0.0
2220	13	0.9	19	1.7	22	0.0	26	0.0	22	1.4	22	1.5
2230	22	0.8	21	0.0	20	0.0	36	3.3	4	0.6	28	0.0
2240	16	3.0	6	3.9	6	1.8	2	4.5	32	4.9	16	0.0
3172	0	0.0	7	0.4	6	0.5	6	0.0	10	1.0	0	0.0
3173	0	0.0	0	0.0	7	0.0	3	0.0	0	0.0	0	0.0
Total	78	[illegible].3	72	6.0	85	2.6	73	10.4	75	10.2	66	1.5

C. Percentages of Questions (Q) and Marks (M)

Teacher	3		8		11		12		13		15	
Topic	Q	M	Q	M	Q	M	Q	M	Q	M	Q	M
2205	29	11	18	0	16	0	0	25	0	22	0	0
2210	5	0	8	0	12	13	0	0	9	0	0	0
2220	17	18	26	28	26	0	36	0	29	14	33	100
2230	28	16	29	0	24	0	49	32	5	6	42	0
2240	21	56	8	65	7	67	3	43	43	48	24	0
3172	0	0	10	6	7	21	8	0	13	10	0	0
3173	0	0	0	0	8	0	4	0	0	0	0	0
Total	100	100	100	100	100	100	100	100	100	100	100	100

Table 10: Weights of Tests, Exams and Other Criteria in Final Grades

	First Half						Second Half						Weighting of Halves
Tchr	Tests		Other			Exam	Tests		Other			Exam	
	N	%	Q	A	P		N	%	Q	A	P		
01n	5	30				20	5	30				20	50:50
01x	5	30				20	5	50					50:50
02	4	25				10	4	25				40	35:65
03	4	23			2.5	25	3	17		6	2.5	25	50:50
04	3	30					3	30				40	30:70
05n	3	15			5	30	3	15			5	30	50:50
05x	3	15			5	30	3	50					50:50
06n	3	13	2	4		20	4	40				20	40:60
06x	3	17	3	6		25	4	50					50:50
07n	3	18				12	3	30				40	30:70
07x	3	30				20	3	50					50:50
08	3	19	7				7	31	4			40	25:75
09	3	18	5				4	25	13			40	23:77
10	4	26				15	6	35				25	40:60
11	6	26		3			7	32				40	30:70
12	5	28				10	5	28		5		30	38:62
13	4	31					5	34				35	30:70
14	5	24				17	4	24				35	40:60
15	3	14	8			9	5	23	4	3		40	30:70
16a	1	15	10		10	15	1	15	10		10	15	50:50
16b	2	30	15		15	15			5		5	15	75:25
17	4	22	1				5	25	3	5	10	35	23:77

Note: Q = Quizzes; A = Assignments; P = Participation Mark

n = not exempted from writ... final exam
x = exempted
a = monthly grade from first half eliminated from final grade calculation
b = monthly grade from second half eliminated

Table 11: Testing Time in Hours

Teacher	Quizzes[1]	Tests	Midterm	Final	Total
1		12.0	2.0	2.0	16.0
2		8.8	1.0	2.0	11.8
3		8.2	1.5	2.0	11.7
4		7.0		2.0	9.0
5		7.6	1.5	2.0	11.1
6	1.3(3)	8.2	2.0	2.0	13.5
7		7.0	1.5	1.5	10.0
8	2.3(7)	9.4		2.0	13.7
9	4.2(10)	8.2		2.5	[illegible].9
10		12.0	1.5	3.0	16.5
11		15.2		2.0	17.2
12		9.5	1.0	2.0	12.5
13		8.9		2.0	10.9
14		10.5	2.0	2.0	14.5
15	2.3(9)	9.3	1.2	2.0	14.8
16[2]	6.0(28)	3.0	1.0	1.0	11.0
17	0.7(3)	9.2		2.0	11.9

[1]Number of quizzes in brackets

[2]Teacher 16 had five monthly one hour tests during the semester. The third and fifth were designated the midterm and final exams, respectively.

Table 12: Lengths of Teaching Segments (in Hours) Prior to Tests and Examinations

Teacher	First Half Test						Mid Term Exam	Second Half Test							Final Exam
	1	2	3	4	5	6		1	2	3	4	5	6	7	
01	9	9	13	11	6		4	8	10	10	6	5			1
02	7	10	13	11			1	10	7	13	10				10
03	11	12	10	10			1	19	20	10					3
04	9	11	12				*	10	14	20					12
05	8	18	16				5	13	4	16					8
06	14	15	14				5	24	7	10	3				2
07	10	17	14				5	23	12	9					6
08	14	9	9				*	9	9	9	9	8	9	8	1
09	13	9	10				*	21	17	13	5				1
10	9	10	8	5			13	6	10	8	17	4	5		1
11	5	5	4	4	5	9	*	5	5	8	5	9	9	6	7
12	4	9	8	6	[illegible]		6	8	12	8	5	7			10
13	6	14	10	11			*	10	14	9	9	9			7
14	10	7	11	8	8		1	9	9	11	11				5
15	13	8	11				5	10	10	9	10	12			5
16[1]	16	16					16	25							25
17	10	9	11	12			*	16	14	8	10	7			4

[1]Teacher 16 gave five monthly tests; the third and fifth were designated as the midterm and final exams.

* midterm exams were not administered by these teachers.

Table 13: Numbers of Items on Term Tests

Teacher	Test													Total	Average/test
	First Half						Second Half								
	1	2	3	4	5	6	1	2	3	4	5	6	7		
01	26	19	9	16	11		15	27	25	25	7			180	18
02	10	9	12	11			7	8	11	15				83	10
03	15	15	13	12			10	9	9					83	12
04	11	8	13				8	6	8					54	9
05	22	27	17				10	10	8					94	16
06	16	22	25				5	20	7	12				107	15
07	15	12	7				7	18	12					71	12
08	14	29	13				11	16	7	5	13	19	5	132	13
09	20	22	13				7	15	32	20				129	18
10	11	13	7	15			8	13	7	20	23	9		126	13
11	11	12	12	10	8	6	4	9	16	7	15	12	7	129	10
12	10	13	10	11	6		6	12	7	9	8			92	9
13	10	6	17	7			7	10	5	15	12			89	10
14	10	16	10	13	5		8	8	14	18				102	11
15	25	15	25				12	10	15	12	11			125	16
16[1]	25	17					11							53	18
17	5	14	4	16			4	15	15	22	25			120	13
											Average:			104	13

[1]Teacher 16 also gave 28 quizzes, totalling six hours testing time.

Table 14: Numbers of Content Categories Covered[1]

Teacher	On Term Tests	On Midterm	On Final	On Final Only
1	14	11	7	
2	13	7	11	
3	12	8	13	1
4	10		9	2
5	11	6	8	3
6	11	6	6	
7	11	7	9	
8	10		10	
9	13		11	
10	11		10	
11	11		9	
12	10	7	10	
13	13		12	
14	10	7	10	
15	11	6	9	
16[2]	9	2	4	3
17	11		11	1

[1]Total number of content categories covered is found by adding the first and last columns.

Table 15: Midterm and Final Grade Averages

Teacher	Midterm Grade	Final Grade	Difference
01[1]	68	67	-1
02	78	73	-5
03	75	67	-8
04	61	54	-8
05[1]	73	73	-0
06[1]	76	72	-4
07[1]	75	69	-6
08	79	64	-15
09	65	60	-5
10	72	64	-8
11	72	70	-2
12	76	74	-2
13	74	65	-9
14	72	65	-7
15	70	62	-8
16	76	75	-1
17	77	65	-12
Average:	73	67	-6

[1]These teachers followed exemption policies; the final grades reported here are for all students, including those exempted.

Table 16: Means and Standard Deviations of Term Marks and Final Exam Marks

Teacher	Term Mark Mean	Term Mark S.D.	Final Exam Mean	Final Exam S.D.
No Exemption Policy				
2	74	14.4	71	22.5
3	73	16.2	50	23.4
4	59	20.9	47	20.8
8	71	23.3	53	26.5
9	63	25.4	56	29.5
10	67	16.0	58	25.0
11	77	16.0	67	19.0
12	75	11.6	70	20.0
13	69	13.3	58	21.2
14	69	18.3	59	23.9
15	68	14.0	54	14.3
16	78	14.8	63	24.7
17	73	13.0	50	19.1
Average	70	16.7	58	22.3
Exemption Policy[1]				
1	58	10.1	40	20.0
5	62	7.9	53	12.3
6	53	8.4	24	19.7
7	55	7.2	36	[illegible].3
Average	57	8.4	38	17.8

[1]Based only on students not exempted

Table 17: Coefficients of Correlation Between Term Tests and Final Examination

Teacher	First Half Tests 1	2	3	4	5	6	Second Half Tests 1	2	3	4	5	6	7	Average Correlation
2	60	63	64	69			50	73	78	67				66
3	02	73	77	73			[illegible]	70	56					61
4	65	67	77				76	84	87					76
8	87	88	79				80	92	87	92	90	90	98	88
9	83	94	93				96	91	97	96	90			93
10	78	69	55	71			71	58	76	89	86	82		74
11	84	75	79	72	68	79	84	83	82	79	79	65	81	78
12	58	54	43	65	61		81	82	55	54	84			64
13	69	73	53	76			57	66	74	85	76			70
14	87	81	74	82	76		86	87	89	88	-			83
15	71	72	71				37	70	67	77	65			66
17	61	69	56	70			61	64	61	73	85			67
													Average:	74

Table 18: Maximum Number of Marks Allocated to Questions and Sub-Questions

	Teacher No.																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Total Marks:	130	119	96	78	103	123	105	121	120	144	125		105	116	123	109	126
Questions																	
1	5	3	3	3	4	5	3	4	3	5	5	4	3	4	4	4	4
2a	2	2	2	1	1	1	1	2	1	1	1	2	1	1	1	2	1
b	2	1	2	1	1	1	1	2	1	1	1	2	1	2	1	2	2
c	6	4	3	3	3	4	3	4	4	6	4	4	4	6	5	5	3
d	3	3	2	2	2	3	2	3	2	3	3	3	2	3	3	2	3
e	2	3	2	2	2	2	2	3	2	2	2	2	2	2	2	2	2
f	5	4	4	2	3	3	3	4	3	5	3	3	3	4	4	3	4
g	2	4	2	1	2	2	2	2	3	2	2	2	2	2	2	2	2
h	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
i	2	3	2	3	1	2	2	2	4	2	2	3	3	3	4	2	2
j	2	3	1	2	2	2	2	3	4	2	2	2	2	3	2	2	2
k	3	3	2	2	3	3	2	3	4	3	3	2	2	3	3	3	3
l	3	3	3	2	2	3	2	3	3	3	2	2	2	3	2	3	3
m	4	3	3	2	3	3	2	3	3	3	2	3	3	2	3	2	2
3a	2	2	2	1	2	2	1	2	1	2	2	2	1	1	2	1	2
b	2	2	2	1	2	2	2	2	2	2	1	2	2	2	2	2	2
c	2	2	2	1	2	2	2	2	2	2	1	2	1	2	1	2	2
d	3	3	3	2	4	3	2	3	3	4	3	2	3	3	4	2	4
e	2	2	2	1	3	2	2	2	2	2	3	2	2	2	2	2	2
f	4	5	5	3	3	7	5	4	6	6	5	4	7	3	7	4	7
g	3	5	4	4	3	6	4	6	5	5	4	4	5	3	7	4	4
h	2	2	2	2	3	3	2	2	3	3	3	2	2	1	2	2	2
i	2	3	2	1	3	3	2	2	3	3	4	3	2	2	2	2	2
4	3	4	3	2	4	4	4	4	4	5	5	3	3	3	4	3	5
5a	3	2	2	1	2	2	2	3	2	3	.	2	2	2	2	3	3
b	1	2	1	1	1	2	1	2	2	1	2	1	1	1	1	1	1
c	2	2	2	2	2	2	2	2	2	2	3	2	2	2	1	2	
6a	3	2	1	1	2	2	2	2	2	4	3	2	2	3	3	3	3
b	2	2	1	1	2	2	2	2	1	2	2	2	2	3	1	2	2
c	4	3	3	2	3	3	3	3	4	6	6	4	3	4	4	3	5
7	8	9	6	4	6	10	11	8	10	12	12	6	6	9	11	6	7
8tot	14	9	7	7	10	13	8	12	9	15	8	8	9	13	10	12	11
8a	8	4	4	4			4	7	3		5		6	5	6	8	
b	3	3	1	1			2	2	2		2		1	5	2	2	
c	3	2	2	2			2	3	4		1		2	3	2	2	
9	9	5	5	4	5	7	6	6	6	9	8	6	6	7	7	6	10
10	6	5	3	4	5	5	6	6	5	6	5	5	5	4	5	5	7
11	10	7	5	5	5	5	7	6	7	10	9	6	7	6	7	6	8

Table 19: Coefficients of Correlation Among Teacher Marking Guides
(N = 35 Questions and Sub-Questions)

Teacher	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	\	86	87	88	90	86	84	91	82	95	81	94	88	93	84	95	90
2		\	89	87	84	91	94	92	95	92	87	89	90	87	94	85	84
3			\	84	83	91	87	88	87	91	82	89	91	82	91	85	88
4				\	83	87	82	92	86	88	76	92	92	85	87	90	85
5					\	90	83	92	82	92	80	87	83	88	81	90	87
6						\	89	94	90	93	83	90	91	89	93	91	87
7							\	88	94	92	93	89	83	84	92	82	87
8								\	87	93	79	91	89	92	89	95	88
9									\	91	89	88	90	84	94	83	84
10										\	91	95	91	94	93	93	93
11											\	87	82	80	88	76	84
12												\	92	90	91	92	92
13													\	83	93	89	92
14														\	85	94	84
15															\	84	87
16																\	88
17																	

Note: Decimal points omitted throughout the table.

Table 20: Coefficients of Correlation Among Teacher Marks
(N = 20 Students)

	Teacher 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	\	85	83	82	86	87	86	85	91	85	91	89	88	92	87	81	89
2		\	94	88	96	96	93	94	91	96	91	95	92	92	91	92	91
3			\	86	92	94	95	91	92	92	90	93	91	91	91	88	87
4				\	93	86	89	94	82	90	89	94	93	87	84	92	92
5					\	93	94	96	91	94	91	96	96	92	91	92	94
6						\	94	93	91	91	92	94	91	93	92	87	88
7							\	94	93	97	94	95	96	96	96	94	92
8								\	89	95	90	95	95	93	91	95	94
9									\	91	92	95	93	92	92	87	92
10										\	91	96	95	93	95	96	94
11											\	96	93	93	95	90	93
12												\	97	94	94	94	96
13													\	91	92	[illegible]	97
14														\	94	88	90
15															\	89	92
16																\	95
17																	

Note: Decimal points omitted throughout the table.

Table 21: Percentage Marks

	Teacher																		
Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17		
1	51	76	67	79	79	73	70	79	59	76	69	76	75	59	66	84	74	1	0
2	52	80	63	72	80	67	70	77	68	79	70	79	74	62	67	83	74	3	0
3	67	75	63	76	77	71	73	75	70	75	74	77	76	67	68	80	74	1	0
4	36	54	45	50	52	47	48	51	46	50	58	58	50	41	50	59	46	0	7
5	72	82	70	83	85	79	77	82	73	80	79	85	79	78	73	83	80	8	0
6	46	64	54	63	66	59	58	64	56	54	64	65	62	54	51	68	59	0	1
7	42	56	54	65	67	55	63	63	53	65	61	65	64	54	54	74	54	0	1
8	62	77	66	73	79	74	71	74	71	74	78	80	79	66	72	83	73	2	0
9	47	64	54	64	61	56	57	64	53	60	62	67	58	55	55	67	56	0	1
10	70	87	77	90	92	80	83	85	79	89	85	92	88	78	80	94	92	13	0
11	49	75	64	79	78	69	69	74	60	70	72	77	69	65	67	74	63	0	1
12	52	71	58	68	66	59	57	69	57	63	65	68	61	55	54	74	60	0	0
13	54	76	63	71	75	72	65	70	59	68	67	75	68	60	57	74	60	0	0
14	52	77	64	71	76	68	63	71	60	69	70	75	69	60	60	78	67	0	0
15	58	83	71	71	77	81	77	78	73	79	77	81	72	72	74	83	72	4	0
16	42	61	42	55	59	52	45	53	44	48	59	58	49	41	48	57	47	0	9
17	52	74	66	73	76	69	69	75	64	69	74	76	71	61	63	80	66	1	0
18	62	82	74	87	83	77	77	79	75	79	81	88	83	66	68	88	81	8	0
19	40	68	60	62	70	63	57	64	63	50	59	68	61	53	54	64	56	0	1
20	67	87	85	79	86	84	86	80	77	87	83	85	81	77	80	85	77	12	0
H	0	6	1	3	5	3	2	3	0	3	3	6	3	0	2	10	3		
F	7	0	2	0	0	1	2	0	2	1	0	0	1	2	1	0	2		

H Number of marks greater than 79% earned by students/awarded by teachers.

F Number of marks less than 50% earned by students/awarded by teachers.

Table 22: Marks Awarded by Selected Teachers to Selected Student Responses

	Question 1		Question 9		Question 10	
	Tchr 1	Tchr 16	Tchr 1	Tchr 16	Tchr 1	Tchr 16
Student 4	2/5	3/4	2/9	4/6	0/6	0/5
Student 10	3/5	4/4	4/9	6/6	6/6	5/5
Student 17	2/5	[illegible]/4	4/9	4/6	1/6	4/5

	Question 7	
	Teacher 13	Teacher 14
Student 3	6/6	1/9
Student 4	3/6	1/9
Student 10	4/6	2/9
Student 15	2/6	5/9
Student 20	4/6	8/9

Table 23: Summary of Item-Total Correlation Coefficients
(N = 17 teachers)[1]

Question	Median	Minimum	Maximum
1	35	-4	58
2b	49	38	61
c	23	-3	45
d[2]	6	-5	45
f	63	-11	82
g	17	-20	41
h	30	-19	42
i	3	-28	30
j	66	46	79
k	43	24	57
l	45	36	59
3a	24	9	37
b	43	8	56
c	58	31	67
d	65	46	80
e	37	30	43
f	60	35	66
g	9	-12	47
h	17	-22	49
i	56	40	78
4	57	26	74
5a[3]	6	-23	49
b	15	-12	29
6a	41	24	57
b	18	-11	45
7	47	27	61
8	44	14	52
9	39	10	57
10	63	48	70
11	39	21	57

[1]For items 2a, 2e, 2m, and 5c, all or most teachers gave all students the same mark; these items are not reported.

[2]One teacher gave all students the same mark for this item; thus, N = 16.

[3]Six teachers gave all students the same mark for this item; thus, N = 11.

Table 24: Results of Generalizability Analyses of Exam Marks
(Students by Teachers by Questions)

Source of Variance	Degrees of Freedom	All Questions (N = 35)	Skills Questions (N = 26)	Other Questions (N = 9)
Students	19	.0809	.0619	.2826
Teachers	16	.0839	.0399	.3894
Questions	34, 25, 8	1.9604	.2977	5.1117
S x T	304	.0039	.0017	.0217
S x Q	646, 475, 152	.6887	.4525	1.2243
T x Q	544, 400, 128	.3884	.2152	.7103
S x T x Q	10336, 7600, 2432	.2951	.1833	.6067

Table 25: Breakdown of the Marks Assigned by Four Teachers to the Exam Responses of Three Students

Teacher	Percentage of Marks for Basic Skill Questions Given to Student No. 4	10	17	Percentage of Marks for Other Questions Given to Student No. 4	10	17
1	46	88	72	25	49	30
13	52	85	79	49	91	60
14	54	89	68	26	66	53
16	63	92	84	53	96	74

FIGURE

Figure 1: TIME MANAGEMENT: TEACHING SEGMENTS AND TESTING INTERVALS FOR 6 TEACHERS

NOTE: M = Midterm Exam
F = Final Exam

Teacher 1

——1 ——2—— 3 —— 4 —— 5 —— M —— 6 —— 7 —— 8 —— 9 —— 10 — F

Teacher 11

—— 1 —— 2 —— 3 —— 4 —— 5 —— 6 —— 7 —— 8 —— 9 —— 10 —— 11 —— 12 —— 13 —— F

Teacher 08

—— 1 —— 2 —— 3 —— 4 —— 5 —— 6 —— 7 —— 8 —— 9 —— 10 — F

Teacher 04

—— 1 —— 2 —— 3 —— 4 —— 5 —— 6 —— F

Teacher 03

—— 1 —— 2 —— 3 —— 4 — M —— 5 —— 6 —— 7 — F

Teacher 06

—— 1 —— 2 —— 3 —— M —— 4 —— 5 —— 6 — 7 — F

APPENDICES

Appendix A

Teacher/Class Code: ____________________

Date: ________________

Topic or Objective: ____

Time — Sequence of Class Activities

Teacher | Student

00
10
20
30
40
50
60
70
80
(minutes)

General Comments:

PLEASE INDICATE ABOVE:

1. Specific problems or examples used for --
   (a) Presenting New Material,
   (b) Review or Further Explanation,
   (c) Homework Assignment.
   **Note: indicate problem by location in text.

2. Method of covering problem solutions --
   (a) Student solutions on the board.
   (b) Model solution worked on board or overhead.
   (c) Oral discussion.

3. Most difficult problems for students.

Appendix B

Taxonomy of Calculus Topis based on the Grade 13 Calculus Guidelines (1972) and two Grade 13 Calculus Textbooks

1000. Limits, Slope, Simple Derivatives

1100. Limits

1110. Limit of a sequence

1111. Find a term(s) given function or nth term
1112. Find a function given terms
1113. Find the limit of a sequence given term(s)
1114. Find the limit of a sequence given the function
1115. Draw and interpret the graph of an infinite sequence
1116. Find difference between limit of a function and value at large n

1120. Sum of a series

1121. Find terms of inf. series given two of: first term, common ratio, sum
1122. Change periodic decimal to infinite series
1123. Find sum of a given number of terms of a series
1124. Find sum of infinite series
1125. Find n so that sum to n terms differs from limit by given amount
1126. Factor or simplify, involving the formula for the sum of a series
1127. Thinking questions, odd series, pattern recognition, etc.
1128. Write series in sigma notation, vice versa
1129. Manipulation of sigma notation, including array questions

1130. Limit of a function

1131. Problems involving irrational numbers as limits of series
1132. Classify fcns as polynomial, rational alg, alg, or non-alg
1133. Find limit of fcns, including factoring out of denominator
1134. Express the derivative as a limit (of k/h as h -> 0)
1135. Verify or prove sum, product or quotient rules for limits
1136. Use sum, product or quotient rules for limits
1137. Working with epsilon-delta definition of limit
1138. Continuous functions

1200. Slope

1210. Slope of a line

1211. Find either linear equation or points on line given the other
1212. Find slopes of secants to continuous curves
1213. Find slopes of tangents by sandwiching method
1214. Find slopes of secants to discontinuous curves
1215. Find slope of tangent by limit method
1216. Questions on graphing curves using tangents by limit method
1217. Review: line lengths, pts of intersection, fcn. notation

1300. Techniques of Differentiation

1310. Rules for differentiation

1311. Find the derivative as a limit
1312. Find the derivative as an algorithm, including simple polynomials (no fractional or negative powers)
1313. Find derivative involving translation and substitution
1314. Find values of derivates in an interval, including graphing
1315. Find derivative using algorithm, polynomials with fractional and negative powers or literal coefficients
1316. Thinking problems not covered above, inc. derivation of the following

1317. Find derivative by application of the chain rule, product rule, and/or quotient rule, with whole number, fractional and/or negative powers

1320. Derivatives of special functions

1321. Find limits of trigonometric functions
1322. Find derivative by algorithm of sine and cosine functions
1323. Graph and/or derive derivatives of complex trig fcns
1324. Find, work with derivatives of reciprocal functions (algorithm)
1325. Working with domains, graphs and derivatives of inverse functions
1326. Understanding properties of the logarithmic function
1327. Find derivatives involving log fcns, fractional and negative exponents and trigonometric fcns included (algorithmic)
1328. Understanding properties of the exponential function
1329. Find derivatives involving exponential functions (algorithmic)

2000. Applications of Differentiation

2100. Equations of Tangents to Curves

2110. Find the equation (or slope) of the tangent to y=f(x)
2120. Find equations (or slopes) for tangents to general curves using implicit differentiation
2130. Finding derivatives of fcns. represented parametrically
2140. L'Hopital's Rule

2200. Rate of Change

2205. Inst. vs. avg. rate of change, related problems
2210. Find velocity from speed, time, using graphs and tables
2220. Find velocity as a derivative
2230. Acceleration: simple graphical and computational practice
2240. Acceleration and closely related problems
2250. Rate problems involving area & volume
2260. Rate problems involving distances
2270. Review: area, volume, plane figures

2300. Curve Sketching, increasing & decreasing fcns

2310. Sketch curves, determine intervals when fcns are increasing or decreasing
2320. Find stationary points for fcns, incl. maxima and minima

2400. Maxima and minima problems

2410. Find sums and differences of numbers
2420. Max/min problems, one-dimensional distance, velocity, acceleration
2430. Find the distance of closest approach
2440. Find area & volume
2450. Find a best angle or a related trigonometric problem
2460. Find best volume, selling price, etc.
2470. Find most efficient electricity or fuel cost
2480. Miscellaneous (eqn given)

2500. Problems involving growth and decay, exponential and logarithmic fcns

2600. Second Derivative

2610. Graph fcn, making use of 2nd derivative, points of inflection, maxima and minima, finding hills and valleys
2620. Algorithmic practice with second derivative
2630. Thinking questions involving 2nd derivative
2640. Finding asymptotes, graphing asymptotic relations
2650. Interpreting graphs

3000. Integration

3100. Functions with a Given Derivative

3110. Find a function with a given derivative, either primitive or subject to an outside condition
3120. Problems related to above
3130. Find primitives and integrals of trigonometric fcns
3140. Find primitives and integrals of exponential and logarithmic fcns
3150. Find families of curves with given slopes

3160. Differential equations

3161. Solve simple differential equations
3162. Problems involving differential equations

3170. Problems involving integration

3171. Simple algorithm practice for motion problems
3172. Motion problems involving quadratic relations
3173. Motion problems involving trigonometric functions
3174. Electrical problems involving trigonometric functions
3175. Growth and decay problems, involving exponential and logarithmic functions

3180. Differentials

3181. Algorithm practice
3182. Problems involving differentials

3200. Definite and Indefinite Integration

3210. Area

3211. Review, including using Archimedes' method for finding area
3212. Find the area function for area swept out by an ordinate segment
3213. Find area bounded by linear function, an axis, and straight lines
3214. Find area function for area under a curve
3215. Find area bounded by nonlinear fcns, axis, and two lines

3220. Indefinite integration

3221. Integrate, using simple rules
3222. Integrate by substitution, parts, or other complex procedure

3230. Integration and area

3231. Find area by method of summation
3232. Find the definite integral
3233. Find the area between two curves
3234. Find the area between exponential or logarithmic curves

3240. Integration and volume

3241. Find the volume of a solid
3242. Find the volume of revolution of simple fcns
3243. Find the volume of revolution of trig fcns
3244. Find the volume of revolution of exponential & logarithmic fcns

3250. Other applications of integration

3251. Problems involving work and pressure
3252. Find the length of a curve
3253. Find approximate values of integrals or areas using the Simpson or trapezoidal rules

3254. Find the area of a surface of revolution
3255. Find the average value of a function

4000. Polar Coordinates & Complex Numbers

4100. Polar Coordinates

4110. Plot points in polar coordinates
4120. Convert from polar to rectangular coordinates and vice versa
4130. Sketch and investigate equations in polar coordinates.
4140. Use polar derivative fcn to find angle between radius vector and tangent
4150. Find areas in polar coordinates
4160. Find the length of an arc

4200. Complex Numbers

4210. Simple familiarization and plotting of points in complex coordinates
4220. Add complex numbers
4230. Multiply complex numbers
4240. Simple familiarization with complex conjugate numbers
4250. Divide complex numbers
4260. Solve quadratic equations using basic formula

4270. Polar form

4271. Convert complex numbers from polar to rectangular and vice versa, plot same
4272. Multiply and divide complex numbers in polar form
4273. Evaluate functions using De Moivre's theorem

4280. Find the roots of complex numbers
4290. Miscellaneous

Appendix C: Content Category Membership of Calculus Topics

Category I.D. Number	Topic I.D. Numbers in the Taxonomy
1	1131 1132 1133 1136 1137 1138
2	1111 1112 1113 1114 1115 1116 1121 1122 1123 1124 1125 1126 1127 1128 1129
3	1312 1313 1314 1315 1317 1321 1322 1323 1324 1325 1326 1327 1328 1329 2620
4	3110 3120 3130 3140 3161 3171 3181 3221 3222 3253
5	1134 1135 1311 1316 2630
6	1211 1212 1213 1214 1215 1216 1217 2110 2120 2130 2140
7	2310 2320 2610 2640 2650 3150
8	3211 3212 3213 3214 3215 3231 3232 3233 3234
9	3241 3242 3243 3244 3252 3254 3255
10	2205 2210 2220 2230 2240 3172 3173
11	2250 2260 2270 2500 3162 3174 3175 3182 3251 3256
12	2410 2420 2430 2440 2450 2460 2470 2480
13	4210 4220 4230 4240 4250 4260 4271 4272 4273 4280 4290
14	4110 4120 4130 4140 4150 4160

Appendix D: Percentage of Homework Questions by Content Category

Teacher	Content Category 1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	5.9	3.7	23.8	15.9	0.1	8.2	3.8	2.9	2.4	2.3	7.4	3.4	13.1	7.2
2	6.9	4.9	30.0	13.8	0.2	4.4	6.7	4.7	0.6	12.8	10.4	1.8	0.6	2.2
3	7.3	4.0	39.9	13.8	0.1	7.5	5.8	4.5	0.9	5.2	4.7	2.2	0.3	3.9
4	3.5	4.6	33.6	11.0	0.2	12.1	8.8	7.9	1.5	6.4	8.6	2.0	0.0	0.0
5	6.4	3.0	52.4	11.7	0.0	6.2	1.3	7.1	1.4	2.5	1.6	1.9	1.1	3.5
6	5.2	0.0	39.1	20.0	0.0	6.3	3.9	7.4	0.2	7.6	1.7	8.5	0.0	0.0
7	8.3	7.0	33.5	17 1	1.4	4.6	4.8	3.5	0.0	9.8	6.5	3.5	0.0	0.0
8	11.0	13.0	38.1	13.4	1.7	4.8	6.2	0.8	0.0	5.2	4.0	1.8	0.0	0.0
9	4.4	2.9	29.2	12.8	0.2	5.1	9.2	8.1	2.6	2.7	4.6	6.4	9.5	2.4
10	0.0	0.0	25.8	14.8	0.2	11.0	13.2	8.4	6.6	10.8	3.4	3.8	0.0	2.1
11	5.9	2.9	28.2	20.0	2.1	7.1	9.5	5.0	0.0	8.8	6.7	3.8	0.0	0.0
12	12.7	0.0	41.6	16.1	0.3	4.7	5.4	5.4	0.0	6.7	4.4	2.6	0.0	0.0
13	7.4	4.7	21.9	12.2	0.0	7.1	6.7	6.1	4.6	7.7	5.3	7.2	8.0	0.0
14	8.1	0.9	29.6	14.9	0.1	7.3	11.6	4.7	0.0	13.2	3.3	6.3	0.0	0.0
15	7.7	0.0	36.6	21.4	0.5	3.8	6.7	8.9	2.9	4.9	3.8	2.8	0.0	0.0
16	7.8	2.4	19.3	20.6	2.1	4.1	9.2	4.7	0.5	5.9	10.0	2.0	6.9	4.6
17	2.1	3.5	34.6	13.8	0.3	10.8	7.7	6.1	1.9	10.0	1.8	7.3	0.1	0.0
Median	6.9	3.0	33.5	14.8	0.2	6.3	6.7	5.5	0.9	6.7	4.6	3.4	0.0	0.0

Appendix E: Test Marks as Assigned per Teacher by Content Category

	Category													
	I				II	III		IV		V			VI	
Teacher	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Main Data Set														
1	4.6	2.0	12.1	16.2	2.5	5.5	8.5	6.0	6.3	2.5	10.9	6.2	6.1	10.5
2	4.8	4.8	22.6	17.5	3.0	6.8	6.1	1.5	1.8	7.2	11.1	1.7	7.6	3.5
3	10.0	3.9	27.6	9.3	6.7	9.6	6.8	2.5	1.7	5.3	5.0	5.0	0.0	6.5
4	6.9	2.2	17.5	6.5	6.8	6.3	12.1	6.5	5.1	7.8	13.6	8.8	0.0	0.0
5	9.0	1.6	11.4	21.0	2.2	3.7	9.0	9.8	5.9	5.1	6.6	8.2	1.9	4.6
6	4.2	0.0	21.2	24.3	2.7	6.2	6.9	7.9	5.5	6.8	4.4	9.9	0.0	0.0
7	5.4	3.5	21.6	16.1	1.6	8.5	9.8	8.0	0.0	12.4	5.9	7.4	0.0	0.0
8	6.6	9.0	22.1	20.4	1.3	11.8	5.9	0.0	0.0	6.0	9.1	7.9	0.0	0.0
9	5.3	0.3	22.0	9.5	2.7	5.4	8.2	7.7	3.8	9.9	3.0	8.6	9.8	4.0
10	0.7	0.0	14.4	17.6	2.3	6.2	17.0	5.7	6.6	12.4	3.5	11.2	0.0	2.3
11	6.0	3.7	14.2	10.9	5.5	11.9	15.4	8.2	0.0	2.6	9.5	12.1	0.0	0.0
12	6.9	0.0	21.4	19.8	7.5	7.7	7.1	8.1	0.0	10.4	6.4	4.7	0.0	0.0
13	3.3	2.6	12.9	7.3	3.5	8.9	10.0	6.8	7.8	10.2	10.2	7.0	9.4	0.0
14	9.4	0.0	18.9	4.4	5.4	10.4	13.4	6.7	0.0	15.6	6.1	9.6	0.0	0.0
15	3.9	0.0	25.5	18.7	5.7	6.3	8.2	9.3	4.2	1.5	6.0	10.8	0.0	0.0
16	9.3	0.4	15.8	12.0	3.5	4.5	10.4	3.8	2.4	3.7	12.1	9.4	6.3	6.3
17	2.2	1.5	20.0	10.8	10.2	7.8	7.6	11.7	2.7	6.8	9.3	9.2	0.0	0.0
Median	5.4	1.6	20.0	16.1	3.5	6.8	9.0	6.8	2.7	6.8	6.6	8.5	0.0	0.0
Data on Exemption Policy														
1 (non)	4.6	2.0	12.1	16.2	2.5	5.5	8.5	6.0	6.3	2.5	10.9	6.2	6.1	10.5
1 (ex)	4.5	2.0	12.4	16.3	1.6	5.5	8.5	4.8	6.4	2.5	13.6	7.1	3.0	11.8
5 (non)	9.0	1.6	11.4	21.0	2.2	3.7	9.0	9.8	5.9	5.1	6.6	8.2	1.9	4.6
5 (ex)	8.4	1.5	10.7	17.4	2.1	3.4	8.4	23.2	6.4	4.8	6.2	7.5	0.0	0.0
6 (non)	4.2	0.0	21.2	24.3	2.7	6.2	6.9	7.9	5.5	6.8	4.4	9.9	0.0	0.0
6 (ex)	5.7	0.0	15.1	26.5	3.7	8.4	4.4	8.4	4.8	5.5	4.0	13.5	0.0	0.0
7 (non)	5.4	3.5	21.6	16.1	1.6	8.5	9.8	8.0	0.0	12.4	5.9	7.4	0.0	0.0
7 (ex)	7.2	5.7	23.3	8.0	2.7	6.9	10.0	8.7	0.0	11.6	9.7	6.0	0.0	0.0
Data on Teacher 16, Three Different Test Dropping Options														
All inc.	9.3	0.4	15.8	12.0	3.5	4.5	10.4	3.8	2.4	3.7	12.1	9.4	6.3	6.3
Drop T1	1.1	0.0	15.4	14.7	1.9	4.2	10.3	4.6	2.9	3.3	14.8	11.4	7.7	7.7
Drop T2	10.4	0.5	3.9	15.0	2.4	1.3	11.5	4.7	3.0	4.6	15.1	11.7	7.8	7.9
Drop T4	12.2	0.6	20.7	5.6	4.6	5.9	12.0	0.0	3.2	4.8	14.0	0.0	8.3	8.3

Appendix F: Number of Content Categories Covered per Test

Teacher	First Half Test						Second Half Test							Average
	1	2	3	4	5	6	1	2	3	4	5	6	7	
01	2	4	3	4	3		2	8	5	1	1			3.3
02	3	3	4	2			2	3	3	1				2.6
03	[illegible]	6	2	4			5	3	1					3.3
04	3	4	3				2	3	3					3.0
05	3	1	3				2	1	2					2.0
06	5	3	4				3	1	2	1				2.7
07	3	5	2				3	4	2					3.2
08	3	6	2				5	3	2	1	1	3	1	2.7
09	4	3	3				3	5	4	2				3.4
10	4	2	1	3			3	2	3	6	5	2		3.1
11	2	3	2	3	2	3	2	4	3	3	5	2	1	2.7
12	2	2	2	3	2		3	2	2	2	1			2.1
13	3	3	3	3			3	4	1	8	1			3.2
14	1	4	2	1	2		1	3	4	5				2.6
15	4	3	4				2	3	1	2	3			2.8
16	6	5					3							4.7
17	2	2	2	2			3	5	5	4	4			3.2

Appendix G: Number of Tests on which Each of 14 Content Categories Appeared

Note: * = covered on final examination but not on class tests
' = covered on class tests but not on final examination

Teacher	Content Category													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
01	2'	1'	3	5	2	3'	3'	3	3	1'	3	2'	1	1
02	2	1	2	2	2	3'	1	1	1	2	1	2'		1
03	3	1	4	1	1	4	2	1	*	2	2	1		1
04	2	1'	3	*	2	3	1	1'	*	3'	1	1		
05	1'	1	1	1	*	1'	1	2	1	1'	1'	1	*	*
06	2'		3	2'	2	1'	3	1	1	1	1'	2'		
07	1	1'	2	2	4'	1	1	1		3	2	1		
08	3	2	4	2	1	5	2			4	2	2		
09	1		3	2	2	3'	2'	2	2	2	1	2	1	1
10			4	4	1'	2	6	2	3	5	2	1		1
11	2	2'	7	2	3	2	3	2		4'	4	4		
12	2		5	3	3	1	1	2		2	1	1		
13	2	1'	4	2	1	4	3	2	2	3	2	2	1	
14	3		3	2	1	3	2	2		4	1	2		
15	2		3	4	3	3'	1	2	1	1'	1	1		
16	2'		2'	1	2'	2'	2'	1'	*		1'	1'	*	*
17	2	*	3	4	3	2	3	4	1'	3	2	2		

Marking Standards Study: A List of Examination Items

1. Find the point on the curve $y = 4x^2 + 2x - 3$ at which there is a tangent with slope $- 14$.

2. Find a simplified expression for $D_x y$ in each of the following:

   a) $y = e^{6x}$

   b) $y = \ln \sqrt{e}$

   c) $4x^2 - 4xy + y^2 = 1$

   d) $y = \sqrt{x^2 + 2x}$

   e) $y = x \ln x$

   f) $y = \dfrac{x - 2}{(x + 1)^6}$

   g) $y = \tan x^2$

   h) $y = \cos^2 x$

   i) $y = \tan^{-1}(2x)$

   j) $y = 10^x$

   k) $y = \log_{10}(x^2 - 1)$

   l) $y = x^2 e^{x^2}$

   m) $y - y^2 = 6x$

3. Integrate where possible.

a) $\int x^{\frac{1}{2}} dx$

b) $\int \sin 2x dx$

c) $\int e^{5x} dx$

d) $\int \frac{x^8 - 3x^7 + 4}{\sqrt{x}} dx$

e) $\int \frac{1}{2 + 3x} dx$

f) $\int \frac{7x^2 - 18x + 9}{x^3 - 3x^2} dx$

g) $\int \frac{5}{x^2 + 6x + 13} dx$

h) $\int \sin^2 x \cos x \, dx$

i) $\int x^{-2} a^{x^{-1}} dx$

4. Integrate by parts $\int x \cos x \, dx$

5. Evaluate the following limits if they exist.

a) $\lim_{x \to 3} \frac{x^2 - 9}{x^2 - 3x}$

b) $\lim_{x \to 3^+} \sqrt{9 - x^2}$

c) $\lim \frac{2x^2 - 3x - 14}{4x^2 - 6x - 10}$

6. The motion of a particle away from a fixed reference point is given by $v = 30t^2 - 20t$, where v is velocity in cm/sec and t is time in seconds.

   a) At what time is the acceleration 0?

   b) Determine whether the velocity at this time is a maximum or a minimum.

   c) If the particle's position is $s = 10$ cm when $t = 0$, what is its position after 10 seconds?

7. Find the area enclosed between $y = 2 \sin x$ and $y = \sin 3x$ for $0 \leq x \leq \pi$.

8. For $y = x^3 + 3x^2$

   a) find the co-ordinates of all extrema, identifying each as maximum or minimum.

   b) find the co-ordinates of all points of inflection.

   c) sketch the curve.

9. A dog chases a cat up a tree. At what rate is the distance between them increasing or decreasing (state which) if the dog is 8 feet from the tree and running towards it at 10 feet per second when the cat is 10 feet up the tree and climbing at 7 feet per second?

10. Prove that the volume of a sphere of radius r is $v = \frac{4}{3}\pi r^3$, by revolving a circle of radius r about a suitable axis.

11. A cylindrical metal can is to have a fixed volume of 500 mL. Find the radius and height of the can if the surface area is to be minimum. $V = \pi r^2 h$, $S = 2\pi r h + 2\pi r^2$.

    Use $\pi = 3.14$

TEACHER MARKING SCHEMES

QUESTION #1

Find the point on the curve $y = 4x^2 + 2x - 3$ at which there is a tangent with slope -14.

Teacher 1: (total 5)

$y = 4x^2 + 2x - 3$ (-1) if missing

$y' = 8x + 2$ (1)

But there is a tangent to $y = 4x^2 + 2x - 3$ with slope -14 (1)

$\therefore 8x + 2 = -14$
$8x = -16$
$x = -2$ (1)

At $x = -2$, $y = 4(-2)^2 + 2(-2) - 3$
$= 16 - 4 - 3$
$= 9$ (1)

$\therefore$ at the point (-2,9) on the curve $y = 4x^2 + 2x - 3$ there is a tangent with slope -14 (1)

Teacher 16: (total 4)

slope anywhere $\frac{dy}{dx} = 8x + 2$ (1)

given slope -14

$\therefore 8x + 2 = -14$ (1)
$8x = -16$
$x = -2$ (1)

$y = 4(-2)^2 + 2(-2) - 3$
$= 16 - 4 - 3$
$= 9$

Point is (-2,9) (1)

Answer all questions in the space provided on the exam paper.

1. Find the point on the curve $y = 4x^2 + 2x - 3$ at which there is a tangent with slope - 14.

STUDENT 4

find (x, y) when $f'(x) = -14$

$f'(x) = 8x + 2$

$-14 = 8x + 2$

$-16 = 8x$

$-2 = x$

$f(-2) = 4(-2)^2$, Term 1, students 4, 10, 17

$= 16 + 4 -$

$= 17$

Teacher 1 2/5
Teacher 16 3/4

∴ the point at which the slope of the function $f(x) = 4x^2 + 2x - 3$ equals -14 is (-2, 17)

STUDENT 10

$y' = 8x + 2$

$-14 = 8x + 2$

$8x = -16$

$x = -2$

$P = (-2, 9)$

Teacher 1 3/5
Teacher 16 4/4

So there is a tangent with slope -14 at P.(-2,9)

STUDENT 17

$y = 4x^2 + 2x - 3$

$y' = 8x + 2$

when $m = -14$

$-14 = 8x + 2$

$8x = -16$

$x = -2$

Teacher 1 2/5
Teacher 16 3.5/4

when $x = -2$

$y = 4(-2)^2 - 2(-2) - 3$

$= 16 + 4 - 3$

$= 18$

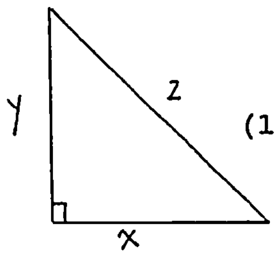
∴ the point on the curve is at (-2, 18)

TEACHER MARKING SCHEMES

QUESTION #9

A dog chases a cat up a tree. At what rate is the distance between them increasing or decreasing (state which) if the dog is 8 feet from the tree and running towards it at 10 feet per second when the cat is 10 feet up the tree and climbing at 7 feet per second.

Teacher 1: (total 9)



(1)

$$\frac{dy}{dt} = 7$$

$$\frac{dx}{dt} = -10$$

$x^2 + y^2 = z^2$ (PT) (1)

$2x \cdot \frac{dy}{dt} + 2y \cdot \frac{dy}{dt} = 2z \cdot \frac{dz}{dt}$ (1)

$\frac{dz}{dt} = \frac{x \cdot \frac{dx}{dt} + y \cdot \frac{dy}{dt}}{z}$ (1)

$\frac{dz}{dt} = \frac{-10x + 7y}{z}$ (1)

When $x = 8$ and $y = 10$,

$z = \sqrt{164}$ (PT)
$= 2\sqrt{41}$ (1)

$\therefore \frac{dz}{dt} = \frac{-80 + 70}{\sqrt{164}}$ (1)

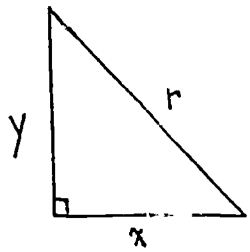
$= \frac{-10}{\sqrt{164}}$

$= -0.78$ (1)

(1)

$\therefore$ when the cat is 10 feet up the tree, the distance between the cat and the dog is decreasing at 0.78 feet per second.

Teacher 16: (total 6)



$\frac{dy}{dt} = 7$ ft/s

$\frac{dx}{dt} = -10$ ft/s

$y = 10,\ x = 8$

$x^2 + y^2 = r^2$

$$2x\frac{dx}{dt} + 2y\frac{dy}{dt} = 2r\frac{dr}{dt}$$

$$x\frac{dx}{dt} + y\frac{dy}{dt} = r\frac{dr}{dt}$$

(1) either of these

$$\frac{dr}{dt} = \frac{x\frac{dx}{dt} + y\frac{dy}{dt}}{r} \quad (1)$$

$$= \frac{\overset{(1)}{8(-10)} + \overset{(1)}{10(7)}}{\sqrt{10^2 + 8^2}} \quad (1)$$

$$= \frac{-80 + 70}{\sqrt{164}}$$

$$= \frac{-10}{\sqrt{164}}$$

$$= \frac{-10\sqrt{164}}{164} \text{ ft/s}$$

or $= \frac{-20\sqrt{41}}{164}$ or $\frac{-10\sqrt{41}}{164}$ or $\approx -.79$ ft/s (1)

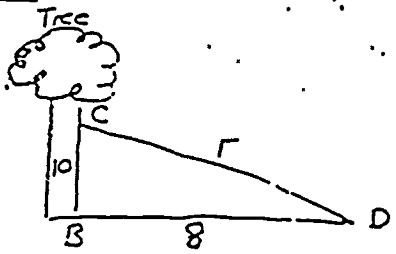
Decreasing by .79 ft/s

or changing by $\frac{-10}{\sqrt{164}}$

Give 4 marks if answer is 11.7 ft/s

9. A dog chases a cat up a tree. At what rate is the distance between them increasing or decreasing (state which) if the dog is 8 feet from the tree and running towards it at 10 feet per second when the cat is 10 feet up the tree and climbing at 7 feet per second?

STUDENT 4



let BD = x ∴ $\frac{dx}{dt}$ = 10 feet/sec
let BC = y ∴ $\frac{dy}{dt}$ = 7 feet/sec

the distance will be decreasing between the dog and the cat ∵ the dog is running faster.

when x = 8 feet + y = 0 feet r = 13 feet
find $\frac{dr}{dt}$

$x^2 + y^2 = r^2$

$2x\frac{dx}{dt} + 2y\frac{dy}{dt} = 2r\frac{dr}{dt}$

$80 + 70 = 13\frac{dr}{dt}$

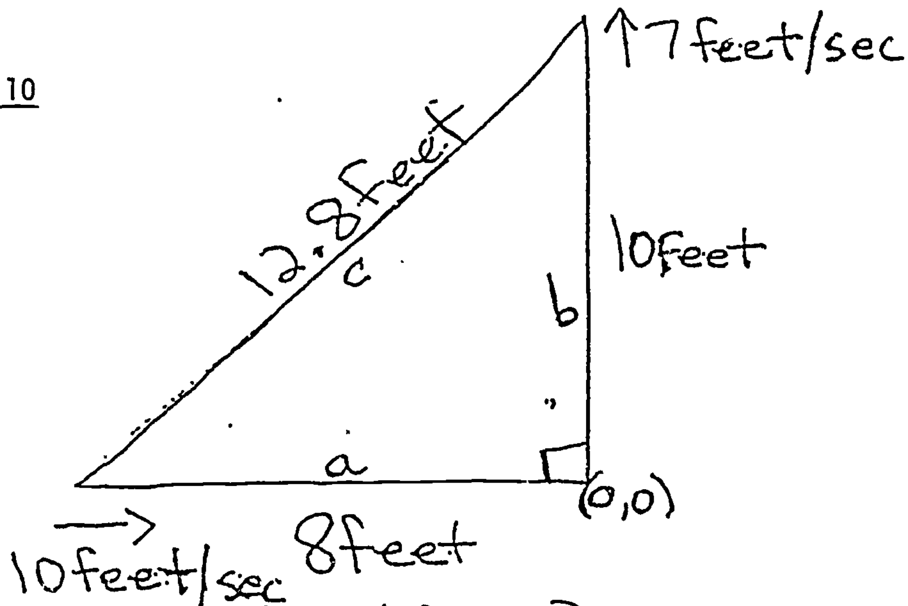
$150 = 13\frac{dr}{dt}$

$11.5 = \frac{dr}{dt}$

Teacher 1 2/9
Teacher 16 4/6

∴ the rate of decreasing in distance between the dog and cat is 11.5 m/sec.

STUDENT 10



$$a^2 + b^2 = c^2$$

$$aa' + bb' = cc'$$

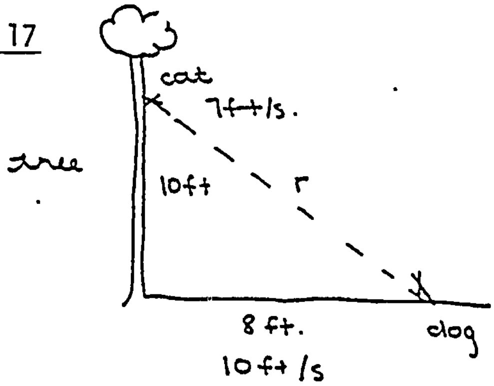
$$(8)(-10) + (10)(7) = (12.8)(c')$$

$$c' = -.78 \text{ feet/sec}$$

∴ the distance is decreasing at .78 feet per second.

Teacher 1 4/9
Teacher 16 6/6

STUDENT 17



Let x be the dog and the distance from the tree

Let y be the cat and the distance up the tree.

by pythagoreum theorem
$x^2 + y^2 = r^2$
$(64) + (100) = r^2$
$r^2 = 164$
$r = \sqrt{164}$

$x = 8ft.$ $\quad y = 10ft$

$\frac{dx}{dt} = 10ft/s.$ $\quad dy/dt = 7ft/s.$

find $\frac{dr}{dt}$

$$x^2 + y^2 = r^2$$

$$x\left(\frac{dx}{dt}\right) + y\left(dy/dt\right) = r\left(\frac{dr}{dt}\right)$$

$$8(10) + 10(7) = \sqrt{164}\left(dr/dt\right)$$

$$150 = \sqrt{164}\left(dr/dt\right)$$

$$\therefore \frac{dr}{dt} = \frac{150}{\sqrt{164}}$$

$$= 11.7 \text{ ft/s.}$$

∴ the rate is increasing (distance) (the cat's getting away!)

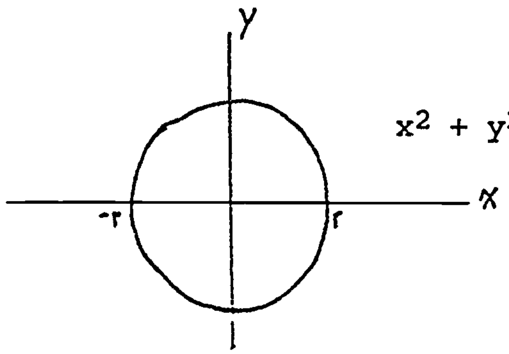
Teacher 1 4/9
Teacher 16 4/6

TEACHER MARKING SCHEMES

QUESTION #10

Prove that the volume of a sphere of radius r is $v = \frac{4}{3}\pi r^3$, by revolving a circle of radius r about a suitable axis.

Teacher 1: (total 6)



$x^2 + y^2 = r^2$ (1)

$$V = \int_{-r}^{r} \pi y^2 \, dx \quad (1)$$

$$= \pi \int_{-r}^{r} (r^2 - x^2) \, dx \quad (1)$$

$$= \pi \left[r^2x - \frac{x^3}{3} \right]_{-r}^{r} \quad (2)$$

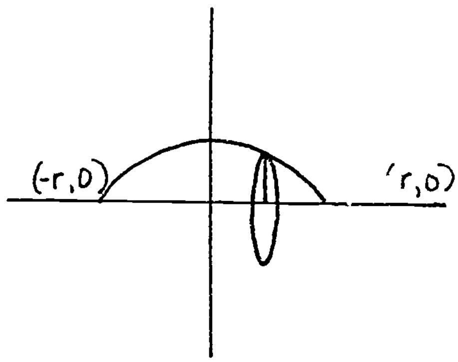
$$= \pi \left[r^3 - \frac{r^3}{3} - (-r^3 + \frac{r^3}{3}) \right]$$

$$= \pi \; (2r^3 - \frac{2r^3}{3})$$

$$= \pi \; (\frac{4r^3}{3})$$

$$= \frac{4\pi r^3}{3} \quad (1)$$

Teacher 16: (total 5)



Circle $\left.\begin{array}{l} x^2 + y^2 = r^2 \\ y^2 = r^2 - x^2 \end{array}\right\}$ (1)

Volume of typical slice is $\pi y^2 \Delta x = \pi(r^2 - x^2)\Delta x$

Volume $= \int_{-r}^{r} \pi(r^2 - x^2)dx$ (1)

(1)

$= \left(\pi r^2 x - \frac{\pi x^3}{3}\right)\Big|_{-r}^{r}$ (1)

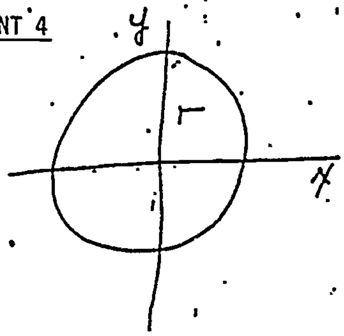
$= \pi r^3 - \frac{\pi r^3}{3} - \left(-\pi r^3 + \frac{\pi r^3}{3}\right)$ (1)

$= 2\pi r^3 - \frac{2}{3}\pi r^3$

$= \frac{4}{3}\pi r^3$

10. Prove that the volume of a sphere of radius r is $v = \frac{4}{3}\pi r^3$, by revolving a circle of radius r about a suitable axis.

STUDENT 4



$$x^2 + y^2 = r^2$$

$$V = \int \pi r^2 dx$$

$$= \int \pi (x^2 + y^2) dx$$

$$= \pi \int x^2 dx + \int y^2 dx$$

$$= \frac{1}{3}\pi x^3 + \frac{1}{3} y^3 + C$$

Teacner 1 0/6
Teacher 16 0/5

STUDENT 10

$$V = \int_{-r}^{r} \pi y^2 dx$$

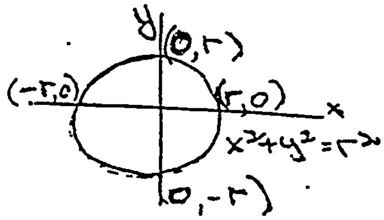
$$= \int_{-r}^{r} \pi (r^2 - x^2) dx$$

$$= \left(\pi r^2 x - \frac{\pi x^3}{3}\right)\Big|_{-r}^{r}$$

$$= \left(\pi r^3 - \frac{\pi}{3} r^3\right) - \left(-\pi r^3 + \frac{\pi}{3} r^3\right)$$

$$= \frac{4}{3}\pi r^3$$

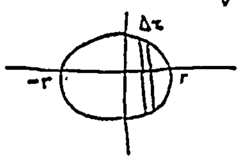
So the volume is $\frac{4}{3}\pi r^3$



Teacher 1 6/6
Teacher 16 5/5

STUDENT 17

equation of circle $= x^2 + y^2 = r^2$



cross-sectional area $= \pi r^2$

$x^2 + y^2 = r^2$ $\quad = \pi y^2$

$y^2 = r^2 - x^2$ $\quad = \pi(r^2 - x^2)$

Teacher 1 1/6
Teacher 16 4/5

$$\text{volume} = \int_{-r}^{r} \pi (r^2 - x^2) dx$$

$$= \pi \left[(r^3/3 - x^3/3) + c \right]_{-r}^{r}$$

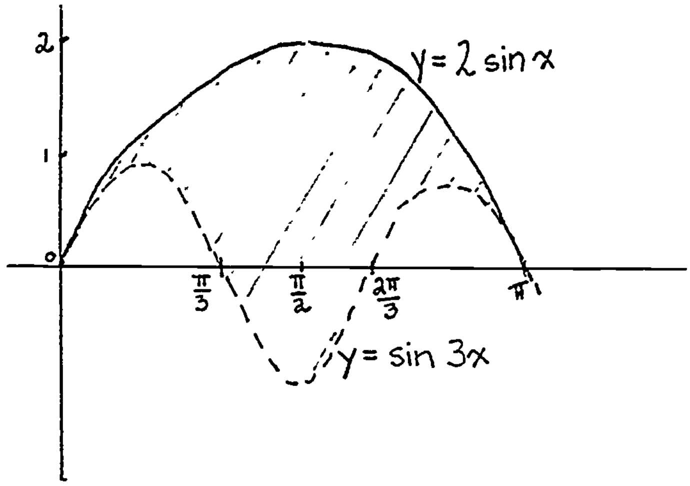
$$= \pi \cdot \left[\frac{r^3}{3} + \frac{r^3}{3} + c\right] - \left[-r^3/3 - r^3/3 + c\right]$$

TEACHER MARKING SCHEMES

QUESTION #7

Find the area enclosed between $y = 2 \sin x$ and $y = \sin 3x$ for $0 \leq x \leq \pi$.

Teacher 13: (total 6)



(2)

$$\Delta A = (y_u - y_2)\Delta x$$

$$= (2 \sin x - \sin 3x)\Delta x$$

$$A = \int_0^{\pi} (2 \sin x - \sin 3x)\,dx \quad (1)$$

$$= \left[-2 \cos x + \frac{1}{3} \cos 3x\right]_0^{\pi} \quad (1)$$

$$= -2 \cos \pi + \frac{1}{3} \cos 3\pi + 2 \cos 0 - \frac{1}{3} \cos 0 \quad (1)$$

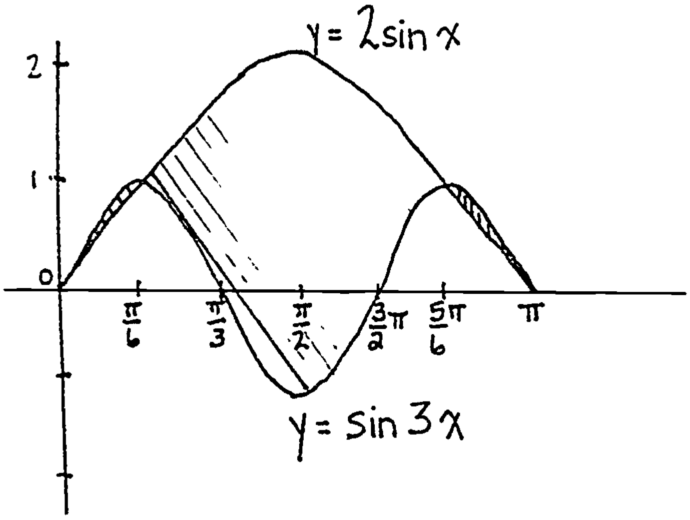
$$= -2(-1) + \frac{1}{3}(-1) + 2\left(\frac{-1}{3}\right)$$

$$= 4 - \frac{2}{3}$$

$$= 3\frac{1}{3} \quad (1)$$

$\therefore$ Area $= 3\frac{1}{3}$ units^2

Teacher 14: (total 9)



(1) sketch with label

(1) correct areas shaded in

$y = 2 \sin x$
$p = 2\pi$

$y = \sin 3x$
$p = \frac{2\pi}{3}$

Points of intersection: $(0°, \pi)$ are obvious

$2 \sin x = \sin 3x$

$2 \sin x = 3 \sin x - 4 \sin^3 x$

$4 \sin^3 x - \sin x = 0$

(1) method $\sin x (4 \sin^2 x - 1) = 0$

$\sin x = 0$ or $\sin^2 x = \frac{1}{4}$

$x = 0, \pi$ $\sin x = \pm \frac{1}{2}$

(1) $x = \frac{\pi}{6}, \frac{5\pi}{6}$

including A = dx

(2) $A = \int_0^{\frac{1}{6}\pi} (\sin 3x - 2 \sin x)dx + \int_{\frac{1}{6}\pi}^{\frac{5}{6}\pi} (2 \sin x - \sin 3x)dx + \int_{\frac{5}{6}\pi}^{\pi} (\sin 3x - 2 \sin x)dx$

(1) integral

$= \left[-\frac{1}{3} \cos 3x + 2 \cos x\right]_0^{\frac{1}{6}\pi} + \left[-2 \cos x + \frac{1}{3} \cos 3x\right]_{\frac{1}{6}\pi}^{\frac{5}{6}\pi} + \left[-\frac{1}{3} \cos 3x + 2 \cos x\right]_{\frac{5}{6}\pi}^{\pi}$

$$= -\frac{1}{3}\cos\frac{\pi}{2} + 2\cos\frac{1}{6}\pi - \left(-\frac{1}{3}\cos 0 + 2\cos 0\right)$$

$$+ \left(-2\cos\frac{5}{6}\pi + \frac{1}{3}\cos\frac{5}{2}\pi\right) - \left(-2\cos\frac{1}{6}\pi + \frac{1}{3}\cos\frac{\pi}{2}\right)$$

(1)

$$+ \left(-\frac{1}{3}\cos 3\pi + 2\cos\pi\right) - \left(-\frac{1}{3}\cos\frac{\pi}{2} + 2\cos\frac{5}{6}\pi\right)$$

$$= -\frac{1}{3}(0) + 2\left(\frac{\sqrt{3}}{2}\right) + \frac{1}{3}(1) - 2(1) - 2\left(\frac{-\sqrt{3}}{2}\right) + \frac{1}{3}(0)$$

$$+ 2\left(\frac{\sqrt{3}}{2}\right) - \frac{1}{3}(0) - \frac{1}{3}(-1) + 2(-1) + \frac{1}{3}(0) - 2\left(\frac{-\sqrt{3}}{2}\right)$$

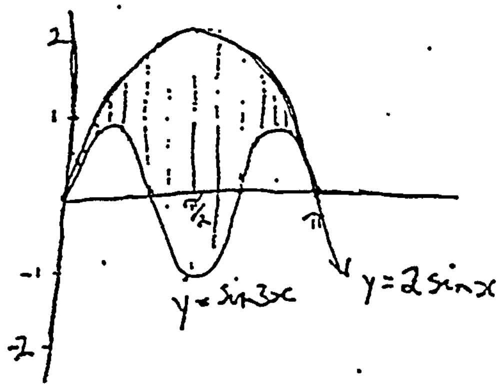
$$= \sqrt{3} + \frac{1}{3} - 2 + \sqrt{3} + \sqrt{3} + \frac{1}{3} - 2 + \sqrt{3}$$

$$= \left(4\sqrt{3} - 3\tfrac{1}{3}\right) \text{ units}^2 \quad (1)$$

$$\doteq 3.6 \text{ units}^2$$

7. Find the area enclosed between $y = 2\sin x$ and $y = \sin 3x$ for $0 \le x \le \pi$

STUDENT 3



$$\text{Area} = \int_{x=0}^{x=\pi} 2\sin x - \sin 3x$$

$$= 2\int_0^{\pi} \sin x - \int_0^{\pi} \sin 3x.$$

$$= \left(-2\cos x\right)\Big|_0^{\pi} + \left(\frac{\cos 3x}{3}\right)\Big|_0^{\pi}$$

$$= \left(2-(-2)\right) + \left(-\frac{1}{3} - \frac{1}{3}\right)$$

$$= 4 - \frac{2}{3}$$

$$= \frac{10}{3}$$

Teacher 13 6/6
Teacher 14 1/9

STUDENT 4.

y = sin 3x is on the bottom
y = 2sin x is on the top

∴ the area between the two functions is

$$A(x) = \int_0^{\pi} 2\sin x - \int_0^{4} \sin 3x$$

$$= \left[-2\cos x\right]_0^{\pi} - \left[-\frac{1}{3}\cos 3x\right]_0^{\pi}$$

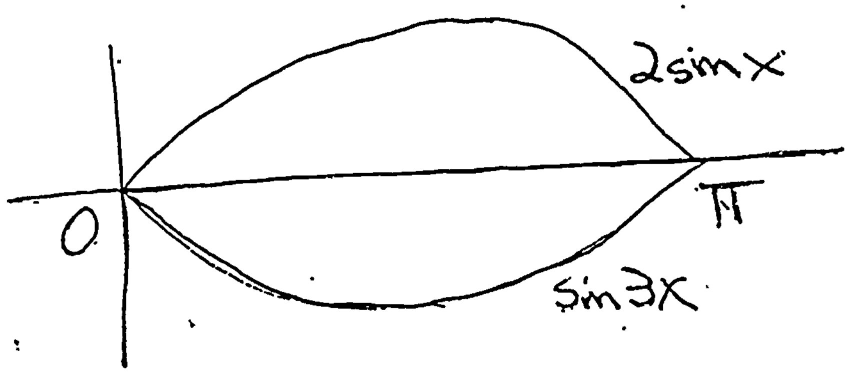
$$= \left| \left[-1.996 - 2\right] - \left[-0.328 - 1\right] \right|$$

$$= \left| -3.996 + 1.328 \right|$$

$$= 2.668$$

Teacher 13 3/6
Teacher 14 1/9

x	0	$\frac{\pi}{4}$	$\frac{\pi}{2}$	$\frac{3\pi}{4}$	π
		1.4	2	1.4	0
$f = 2\sin x$	0				
$g = \sin 3x$	0	.71	-1	.71	0



$$A = \int_0^{\pi} 2\sin x - \sin 3x \, dx$$

$$= \left(-2\cos x + \frac{\cos 3x}{3}\right)\Big|_0^{\pi}$$

[illegible]

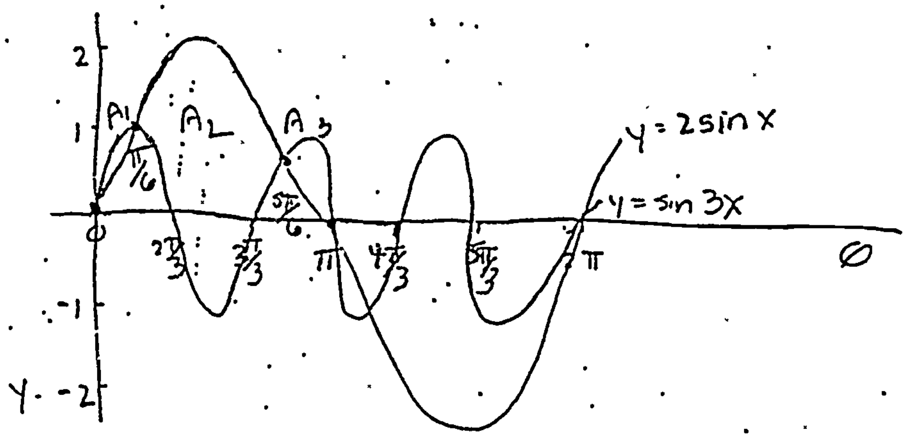
$$= \left(2 - \tfrac{1}{3}\right) - \left(-2 + \tfrac{1}{3}\right)$$

$$= \tfrac{5}{3} + \tfrac{5}{3}$$

$$= \tfrac{10}{3} \text{ sq. units}$$

Teacher 13 4/6
Teacher 14 2/9

STUDENT 15



y = 2 sin x ∴ amplitude = 2, period = 2π

y = sin 3x ∴ period = 2π/3

to find points of intersection:

$\sin 3x = 2\sin x$

~~0 = 2sin x − sin 3x~~

$\frac{\sin 3x}{2} = \sin x$

$x = 30° = \pi/6$

$x = 150° = 5\pi/6$

$A_1 = \int_0^{\pi/6} (\sin 3x - 2\sin x)\,dx$

$= \left[\frac{1}{3}\cos 3x + 2\cos x + c\right]_0^{\pi/6}$

$\doteq 1 + 2 + ¢ - [0 + 1.7 + ¢]$

$\doteq 1.26$

$A_2 = \int_{\pi/6}^{5\pi/6} (2\sin x - \sin 3x)\,dx$

$= \left[2\cos x + \frac{1}{3}\cos 3x + c\right]_{\pi/6}^{5\pi/6}$

$\doteq 1.73 + 0 + ¢ - [1.73 + 0 + ¢]$

$\doteq 3.46$

$A_3 = \int_{5\pi/6}^{\pi} (\sin 3x - 2\sin x)\,dx$

$\doteq \left[\frac{1}{3}\cos 3x - 2\cos x + c\right]_{5\pi/6}^{\pi}$

$= 0 + (-1.7) + ¢ - [-\frac{1}{3} + (-2) + ¢]$

$\doteq -1.7 + \frac{1}{2} + 2$

$= 0.63$

TOTAL AREA

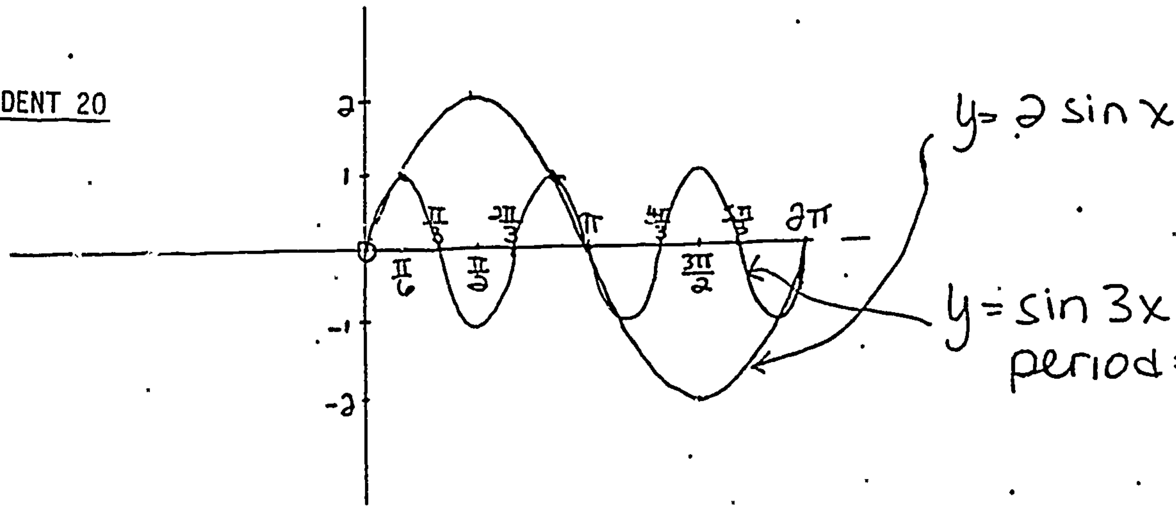
$= A_1 + A_2 + A_3$

$\doteq 1.26 + 3.46 + 0.63$

$\doteq 5.35$

Teacher 13 2/6
Teacher 14 5/9

STUDENT 20



intersection points : $\frac{\pi}{6}$, $\frac{5\pi}{6}$, 0 , π
(found by using calculator)

$A_{①}$ from 0 to $\frac{\pi}{6}$ $= \int_0^{30°} -(2\sin x - \sin 3x)\,dx$

$= -\left[-2\cos x + \frac{1}{3}\cos 3x + c\right]_0^{30°}$

$\doteq -\left[[-1.73 + 0 + c] - [-2 + .33 + c]\right] = .06$

$A_{②}$ from $\frac{\pi}{6}$ to $\frac{5\pi}{6}$ $= \left[-2\cos x + \frac{1}{3}\cos 3x + c\right]_{30°}^{150°}$

$\doteq (1.73 + 0 + c) - (-1.73 + 0 + c)$

$\doteq 3.46$

$A_{③}$ from $\frac{5\pi}{6}$ to π $= -\left[-2\cos x + \frac{1}{3}\cos 3x + c\right]_{150°}^{180°}$

$\doteq -\left[(2 - .33 + c) - (1.73 + 0 + c)\right]$

$\doteq 0.06$

Teacher 13 4/6
Teacher 14 8/9

$\therefore$ Total Area = 0.06 + 3.46 + 0.06
= 3.58 unit ^{2}