National Library of Canada

Acquisitions and Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada

Direction des acquisitions et des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

NOTICE

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

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

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

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

AVIS

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

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

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

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

Canada

University of Alberta

Alberta Diploma Examinations:
Are They Good Predictors of University Success?

by

Kevin W. Twa 

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Education

in

School Psychology

Department of Educational Psychology

Edmonton, Alberta

Fall 1995

ISBN 0-612-06426-3

Canada

University of Alberta

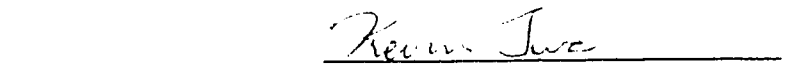
Library Release Form

Name of Author: Kevin W. Twa

Title of Thesis: Alberta Diploma Examinations: Are They Good Predictors of University Success?

Degree: Master of Education

Year this Degree Granted: 1995



Kevin Twa
5801-59 St.
Red Deer, Alberta
T5K 0S5

October 6, 1995

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled Alberta Diploma Examinations: Are they good predictors of university success? submitted by Kevin W. Twa in partial fulfillment of the requirements for the degree of Master of Education in School Psychology.

Dr. Robert Frender

Dr. W. Todd Rogers

Dr. Anne Marie Decore

October 4, 1995

ABSTRACT

The purpose of this study was to compare the predictive validities of diploma examination marks and teacher-assigned high school marks. University GPAs, as well as teacher-assigned marks and diploma examination marks in six core high school subjects, were obtained for the 3,679 undergraduate students enrolled in the Faculty of Education at the University of Alberta in 1993-94. Both teacher-assigned marks and diploma examination marks were moderately correlated with university GPA (0.23 to 0.40). The correlations between teacher-assigned marks and diploma examination marks for the same high school subject were stronger (0.54 to 0.71). Of the several a priori models that were compared, the current weighting of 50% for teacher-assigned marks and 50% for diploma examination marks was the best predictor of university GPA, and it worked as well as an empirically derived multiple regression model. English 30 accounted for more of the variance in the prediction of university GPA than any other high school subject investigated and Math 30 accounted for the least. The combination of English 30, Social Studies 30, and Math 30 scores is proposed as a reasonably effective model to predict university GPA from high school scores.

ACKNOWLEDGEMENTS

At more than one point while I was working on this thesis, I was uncertain if I would finish it. Much credit for its completion goes to Dr. Robert Frender who spent a good deal of time helping me organize my thoughts on this topic. His encouragement and insight helped me to stay focused and aim for attainable goals along the way.

I would also like to thank my other committee members Dr. Todd Rogers and Dr. Anne Marie Decore. Dr. Decore helped me become aware of some important pitfalls in approaching this topic. Dr. Rogers offered valuable, positive criticism to keep me mindful of the possible applications of this research.

Paul Kam, who does computer programming for the Faculty of Education, provided me with the matrix of raw data from which other variables were derived. Jacqueline Leighton and Roy Jaffray from CRAME were especially helpful regarding the multiple regression and cross-validation elements of this thesis.

I would also like to thank Dr. Peter Calder who initially led me to consider Alberta Diploma Examinations as a thesis topic several years ago.

TABLE OF CONTENTS

LIST OF TABLES

Page

CHAPTER 1

INTRODUCTION

I was born in Alberta but I finished high school in British Columbia. Although I wrote scholarship examinations in several subjects at the end of Grade 12, none of those examinations contributed to my final high school grades. Some years later, I returned to attend the University of Alberta. By this time, the Alberta diploma examinations had been recently reinstituted and I heard comments about them from my fellow students. In the fourth year of my education degree, I worked on a research project about the impact of diploma examinations on the teaching-learning process. As I became aware of the differing opinions on the diploma examinations, I was intrigued by the issue that the diploma examinations were motivated more by political than by educational concerns.

For most Grade 12 students the final weeks of June are filled with anticipation. It is a time for graduation, a time for getting out into the real world, a time for thinking about a summer or permanent job, and a time for diploma examinations. For those students who plan to attend a college or university, the diploma examinations are important because half of students' final marks will come from their diploma examination results in their core high school subjects.

Many students will also have noticed in the final weeks of their diploma examination courses that their teachers are preparing them to write the exams. This probably means some practice exam writing, some course review, and a few bars of "remember this... it will probably be on the exam." This could lead students to wonder how it is that they find themselves in this testing situation. Because most high school students do not go on to attend a post-secondary college or university, it might seem a waste of time to study too hard for the last exams they ever hope to face.

For those students planning to continue their education, diploma examination scores will play an important role. For the core subjects, including English 30, Social Studies 30, Mathematics 30, and the sciences, the diploma examination mark in that subject will count for as much as the rest of the work combined. Because of the weight of these examinations, they partly affect whether or not the student will have good enough grades to qualify to enter post-secondary programs.

Although the diploma examination is a high school leaving examination, and not an admissions test, it has important ramifications in that regard. It is worth comparing the diploma examinations to the Scholastic Aptitude Test (SAT), which is used to help make admissions decisions, because there are some similarities between the two and these will be mentioned in the next chapter. The SAT has received considerable

attention regarding how colleges and universities might best utilize the information derived from student scores to facilitate the selection process.

The diploma exams are not equivalent to minimal competency tests, but they have some similarities with them; both are designed to measure a level of skill that is deemed to be "basic." However, whereas minimal competency tests usually focus on identifying if a student is literate, diploma examinations are designed to identify if the student is competent with the curriculum. Another difference is the type of student that the tests are designed to evaluate. Minimal competency tests differentiate literate from illiterate students, whereas diploma exams differentiate among students across the whole spectrum of ability.

Prior to 1973 departmental examinations counted for one hundred percent of a student's grade in that subject. In the period between 1973 and 1983, teacher evaluations of students counted for one hundred percent of the student's grade. When the diploma examinations became mandatory again in 1983, the student's final grade was based on an averaging of the teacher awarded mark and the diploma examination score. This effectively is a weighting of fifty percent from each component. The literature does not suggest an educational reason for selecting a 50-50 weighting.

Research conducted by Calder (1990a) and sponsored by the Alberta Teachers' Association found that 71% of students felt that diploma examinations should be weighted less than the current 50% weighting toward their final grade. Differences existed between the weighting of teacher-evaluated work and diploma examination marks preferred by students and the weighting preferred by teachers (Calder, 1990a). A majority of 64% of students desired a 75/25 split favoring teacher-assigned marks, whereas only 25% favored the current 50/50 split. Among diploma exam teachers there was an opposite trend, with 64% preferring the current 50/50 weighting and 22% preferring a 75/25 split.

When students were asked to compare which was a better indicator of how well they knew the course content, 63% replied teacher-assigned marks were and 24% replied that diploma exam marks were. Among diploma exam teachers, only 6% saw diploma exam marks as more valid than the mark they assigned. Furthermore, Calder found that 60% of diploma exam subject teachers believed that the diploma examinations were required for political rather than educational reasons.

The little research that is available suggests that teacher-assigned marks correlate more highly than diploma examination marks with university GPAs (Mehra, 1977). Based on the work of Mehra, MACOSA (1979, p. 12) reports, "teacher-assigned marks provide as good or better

predictions of later academic success in post-secondary studies than do marks from common examinations." Mehra compared the earlier departmental examinations that were weighted at 100% with the teacher-assigned grades that were weighted at 100% for the period following the discontinuation of the departmental examinations. It is unclear if this finding would be true for the Faculty of Education or for the current system where diploma examinations and teacher-assigned grades are each weighted at 50%.

Purpose of the Study

The current weighting of fifty percent each for teacher marks and diploma exam marks would indicate a belief that each is equally predictive or at least a tacit acceptance of a political compromise to weight these components equally. However, there appears to be neither an educational nor empirical basis for this weighting arrangement. In fact, the literature suggests that politics played a greater role than educational concerns in arriving at a 50/50 split. This raises the issue of whether or not these two methods of evaluation (teacher grade and diploma exam mark) are equally valid assessors of student achievement. The major purpose of my research was to investigate if the current weighting of student grades with fifty percent from teacher evaluations and fifty percent from diploma

examinations is the optimal weighting that can be used when considering student applications to the Faculty of Education.

The present study represents an attempt to identify the predictive validity of teacher marks, diploma examination marks, and a pooling of the two in predicting GPA for students in the Faculty of Education. I hypothesized that among students applying to the Faculty of Education, teacher marks would be more valid predictors than diploma exams of their university GPAs. There were two main reasons for this hypothesis: (a) Mehra's research has shown that teacher evaluations are more highly correlated with GPA than diploma examinations and (b) especially in the Faculty of Education, students are assessed on a range of skills and tasks more like those assessed by the teacher mark than by the diploma examination mark. However, I expected that a pooling of the two ought to be more predictive than either one alone. Although each was apt to contribute, a fifty percent weighting from each was not expected to be the optimal weighting.

Delimitations of the Study

This research was not intended to suggest that decisions should be reduced to statistics alone or that an optimum numeric weighting of certain courses for certain programs would be the solution to deciding which

students should enter the Faculty of Education and which should not. Neither was my research intended to evaluate other criteria such as student enthusiasm, student interest, and student goals that are likely to play a role in student GPA. It was not the goal of this research to propose complicated regression formulas to be applied in some cases where results may demonstrate differences between gender, program type, and grades in a particular course.

It is not possible to take into account such factors as whether a student was performing to the best of their ability or whether certain outside stressors may play a role for some groups more than others.

Significance for the Study

I anticipated that this research would focus attention on any limitations of the current 50/50 weighting of diploma exams with teacher marks that might be found. Depending on results, there may be sufficient cause to investigate the predictive validities of teacher-assigned marks and diploma examination marks of students in other faculties. Further, for teachers, I hoped that some perspective would be brought to the question of whether their evaluations are as valid as the diploma examinations for the purposes of students pursuing entrance to the Faculty of Education.

CHAPTER 2

LITERATURE REVIEW

In this literature review, I begin by providing background information that will explain the reinstitution of the diploma examinations in Alberta. Following this background, the discussion will address general issues in the prediction of college success. After considering a variety of college predictors, the focus will return to the Alberta context to specifically consider the quality of prediction of university success that can be achieved using diploma examination marks and teacher marks.

An Explanation of the Diploma Examinations

In the sections that follow, the historical, political, and definitional issues of the diploma examinations will be explained. Parts of the Historical Issues section that follows are similar to the historical information I wrote for an independent study project when I was an undergraduate. Much of this was incorporated into another project entitled Impact of External Examinations on the Teaching-Learning Process: A Review of the Literature (Calder, 1990b). That review of the literature was part of a study commissioned by the Alberta Teachers' Association to determine the effects of the diploma examinations on the teaching-learning process. The

issues raised during my previous research on that project prompted me to consider how diploma examinations are used for student admissions to the university.

Historical Issues

This look at the historical issues is divided into three periods. The first segment covers a sixty-year period when final marks were entirely decided by a government examination. The second segment considers a ten-year period when final marks were entirely decided by teachers. The last segment considers the present when the teacher mark and the diploma examination mark jointly determine the student grade.

Departmental examinations. For sixty years until 1973, departmental examinations were the sole basis for a student's grades (Symyrozum & Hrabi, 1984). During this period, the provincial government determined which students would pass or fail entirely through the student's performance on one-shot paper and pencil tests. Although students may have suffered from tremendous anxiety because poor performance on a single exam could nullify a whole year of work, "no one questioned the reliability of the results because there was an assumption that the exams were administered strictly and marked fairly and objectively. Teachers did not suggest that it was unfair and unreliable for their students to be tested externally" (Buckley, 1983).

Teacher evaluations. However times changed and the departmental examinations were abolished in Alberta at a time when similar examinations were being phased out across Canada. The role of student evaluation fell to the various school jurisdictions (Anderson, 1983). Killough (1983) indicates that one reason for the discontinuation of the departmentals was improved confidence in teachers ability to evaluate student performance. During this period, the marks issued by teachers formed one hundred percent of a student's final grade. The evaluation by teachers tended to reflect the performance of students over the length of their whole high school course and in a number of different kinds of evaluation beyond paper and pencil tests.

When the pendulum began to swing back toward external evaluations rather than classroom-based evaluation, there was resistance among teachers to external testing. The teachers indicated that the issues of individualism and creativity emphasized in the curriculum were not consistent with external evaluation (Nemeth & Samiroden, 1990). In part, the move toward external evaluations was meant to counteract the growing public perception that the quality of education in Alberta was slipping (Muir, 1983). Between 1973 and 1977, there was a gradual upward "creep" in the distribution of student grades. "Grade inflation" was not unique to Alberta but was widespread through North America

beginning in the 1960s (Huntley, 1976). Three explanations were put forward to explain this trend. First, grade inflation could be due to teachers giving higher and higher grades. Second, it could be because teachers were marking more accurately than the evaluation that was previously done by the departmental examinations. Because the teachers could more carefully consider a student's achievement, it was possible that the student might be more able than a "one shot" examination could reveal. Third, innovation and progress in teaching methods and the quality of the teachers themselves might account for improvement in student grades (MACOSA, 1977).

There was growing concern among the public that student grades might not accurately reflect student achievement. In the absence of standard examinations, perhaps the quality of education was slipping in Alberta. To investigate this concern the Minister's Advisory Committee on Student Achievement (MACOSA) was established in 1976 (Anderson, 1983). Its mandate was to "determine whether the change to non-compulsory Grade 12 departmental examinations has affected the quality of education in Alberta" (MACOSA, 1977, p. 6).

In the first few years of its existence, MACOSA commissioned eighteen studies. These studies sought to accomplish three objectives: (a) review possible changes in the distribution of grades, (b) survey school

policies on student evaluation, and (c) sample various groups' attitudes toward the discontinuation of departmentals.

Regarding the issue of changes in grade distribution, five trends were observed that could explain this situation:

1. In the six matriculation subjects, students scoring "B" or better rose between one-quarter and one-third.

2. One-third fewer students appealed their marks.

3. More "A" and "B" grades were assigned in most courses.

4. Fewer students scored less than 50% than under the departmental system.

5. The percentage of teachers in Alberta with teaching degrees increased approximately 50% between 1969 and 1976 and this improvement in teacher training might explain the improvement in student scores (MACOSA, 1977).

Regarding student evaluation policies, MACOSA (1977) found that school boards differed in the details of their policies, although three general themes seemed to be involved in board policies: (a) objectivity -- evaluation should be free from prejudice and bias; (b) equitableness -- evaluation should dispense equal treatment to all students; and (c) justice -- evaluation should be conducted according to a standard of what was right and proper.

Regarding opinion about the discontinuation of the departmental examinations, MACOSA (1977) conducted a survey of ten distinct groups: school trustees, school superintendents, education consultants, post-secondary school personnel, high school principals, high school teachers, grade 12 students, the public-at-large, public sector employees, and private sector employees. Of the 6,800 surveys mailed out, 4,476 (65.8%) were returned.

Of the many observations that could be made about the results of MACOSA's survey (1977), selected ones follow:

1. Although 17% of respondents felt that the withdrawal of departmentals was positive, 42% thought that it had a negative effect on the quality of education.

2. Between 70% and 80% of each group agreed with the statement "it is difficult to find out how well students in Alberta are achieving" since departmentals were withdrawn.

3. In many groups, over 90% agreed with "an A in one school does not mean the same as an A in another school".

4. The public identified three reasons against reinstating the departmentals: (a) evaluation of students should be based on the year's work; (b) non-examination subjects are not considered to be as important

as examination subjects; and (c) the multiple choice format did not allow testing all of the different kinds of student achievement.

5. Six reasons were identified for reinstating the departmental examinations: (a) examinations had the effect of keeping instruction in each course about the same across the province; (b) examinations meant that all students in the province were judged on the same basis; (c) the examinations certified achievement at the particular standard for the grade twelve level; (d) students learned to work under pressure; (e) students were more likely to work hard; and (f) provincial standards were maintained.

After much study, in 1977, MACOSA recommended that departmental examinations not be reinstated and in 1979 they reiterated that recommendation. Among the factors influencing MACOSA in their recommendation against reinstatement was strong teacher reaction against the examinations on several grounds. Teachers felt that evaluation was a professional responsibility, that one single external examination was an unacceptable mode of evaluation, and that departmental examinations had undesirable effects on the learning environment (MACOSA, 1980). Neither a majority of students nor a majority of teachers were in favor of reinstating departmental examinations. Public opinion was mixed but tended to support the return

of some form of examination that would deal with a perceived erosion of the standards of public education.

Diploma examinations and teacher marks. Diploma examinations were introduced on an optional basis in late 1980 and became mandatory in 1983. In their present form, diploma examinations differed from the previous departmental examinations in that they would count for 50% of the grade rather than the previous 100%. Teachers would be able to contribute to the process of student evaluation, as they had during the years when the exams were not in use, but now as fifty percent of the final student grade.

Then Minister of Education, David King (Minister announces new student evaluation policy, 1983) explained two reasons for the return to standardized testing: (a) to develop and maintain excellence in educational standards throughout the province and (b) to use external examinations as an essential and educationally sound means of testing and monitoring student achievement. In an interview with Jamha (1988), King stated that he considered the reinstatement of diploma examinations to be one of the greatest achievements during his term in office. In the same interview, he continued to say that due to some concerns about how the diploma examinations were being conducted he might consider repudiating his endorsement of them. King was concerned that the diploma examinations

had not resulted in the development of item banks to help teachers draw upon thousands of quality questions. He was also concerned that the diploma examinations did not seem to be developing into a better tool over time. He had expected the examinations to be transitional as better methods were developed beyond multiple choice and short answer questions.

Historically, it would seem that external evaluation is an educational issue subject to pendulum-like trends. For sixty years, external examinations formed the sole basis of student evaluation. Then for ten years, external examinations formed no part in student evaluation. The pendulum did not swing back entirely as the current situation has final marks composed equally by teachers and external evaluations. The decision to settle on a joint method of evaluation was not easily arrived at though.

Political Issues

The political nature of reinstating the diploma examinations becomes clearer as one sees why the government brought them back. Since they did not heed the recommendations of MACOSA or of educators, the explanation must be elsewhere. Ferguson (1983) states that the government commissioned a Gallup poll and the majority were in favor of province-wide achievement tests, although teachers, trustees, and

school associations were opposed to the idea. Ray Martin (1983), an opposition member in the legislature, reported that King stated in the provincial assembly that Gallup polls helped the government to decide to restore comprehensive examinations, but Martin objected to this decision because "government by Gallup polls is not acceptable" (p. 27).

Statements in favor of the examinations. However, the government was not led by polls alone. Public opinion in favor of the examinations found a voice in Ted Byfield (1983), the editor of Alberta Report. He argued that the diploma examinations would establish a level of performance that teachers could work to attain and provide a measure of the quality of teaching being done. There are several problems with his statements though. First, the examinations make no claim to exist to evaluate teachers or schools. Second, because the examinations are norm-referenced with a predetermined mean and no comparability from year to year, at best, the picture that would be provided would be one of stagnation. Hart (1988) considered the possible use of examinations as a means for administrators and school boards to evaluate teachers.

King also expected that the examinations would result in "item banks with thousands of good questions" that teachers could use to improve their teaching (Jamha, 1988). G. Martin (1987) anticipated the development of test items that would be valuable because they were

developed in Alberta on the Alberta curriculum. G. Martin (1983) also suggested that because of (a) improvements in methods of evaluation, (b) a strong involvement of teachers in the examination process, and (c) revisions in the curriculum, the new diploma examinations would be vastly different from the old departmentals. Heck (1983) considered that a beneficial by-product of test-taking was the development of moral values and positive attitudes.

Statements against of the examinations. Anderson (1983) described the difference between the political and educational approach to evaluation. According to her, the political approach intends to make statements about schools, student, systems, teachers, et cetera, while the educational approach aims at affecting the instructional process to the direct benefit of those involved in it. The idea that the diploma examinations exist for political reasons finds agreement with Tomlinson (1984) who considered that diploma examinations are a placebo created to placate the public.

Others did not consider the examinations to be merely a placebo but actually a threat. When the reinstatement of diploma examinations was announced, the Alberta Teachers' Association was originally opposed to the idea. However, they changed their policy on the issue and supported the exams after a time. Some teachers felt that the return to an

external evaluation was a slight against their professional judgment. Anderson (1983; Tomlinson, 1983) reflected this sentiment when she stated that the most effective evaluator of student performance is the student's teacher. In making this remark, she had in mind The Secondary Education Review Project's quote of the Ontario Ministry of Education statement,

> Procedures for evaluating student progress should be sufficiently varied to meet the requirements of different individuals and groups of students, different courses, the several levels of difficulty, and a variety of learning environments...the most effective form of evaluation is application of the teacher's professional judgment to a wide range of information gathered through observation and assessment. (Anderson, 1983, p.16)

The suggestion that the examinations could become a method to evaluate teachers received strong opposition. G. Martin (1987) felt this possibility was "unsettling" (p. 25) and Nemeth and Samiroden (1990) considered the practice "questionable" (p. 28).

Among the detractors of the return to diploma examinations are Nemeth & Samiroden (1990). They identified nine problems with diploma examinations:

1. Diploma examinations contradict the belief in continuous formative evaluation.

2. Since diploma examinations are normed, there is the problem that students do not have the same classroom experiences.

3. Diploma examinations guide the curriculum and cause teachers to teach to the exams (also Stuart, 1987).

4. Diploma examinations substitute content for context.

5. Cognitive level is designed into the exam questions but the examinee determines the kind of thinking that actually is used to complete the question (also Stuart, 1987).

6. Only the limited portion of the curriculum that can be tested as facts are assessed by the exams.

7. Diploma exams are not comparable from test administration to test administration because the questions are revised, some new questions are added, and the students differ from year to year (also Martin, 1987).

8. Comparisons between individuals and jurisdictions are impossible because students with the same score did not answer the same questions correctly.

9. Test results only correlate with future results on similar tests; they are not accurate predictors of future success.

Because the diploma examinations are a kind of certifying examination they have the risk of controlling the curriculum through their heavy use of recall of factual material (Airasian, Madaus, & Pedulla, 1979).

The debate over the reinstatement of a province-wide external examinations brought a number of issues forward both in favor and opposed to it. The arguments in favor tended to be of the kind that the public needs to be reassured that Alberta has a quality system of education and that a standard exam across the province is a good method to achieve that while also providing a good means to compare students from different schools. The arguments against the examination tended to be of two kinds. Some focused on teachers as independent professionals and others identified questions about the ability of an external examination to accomplish educationally sound goals.

In the end, a compromise was reached that permitted teachers to continue to have a role in evaluation while the diploma examinations permitted comparisons of students and an improved public confidence in the quality of education. The type of examination that resulted has some distinct differences from other methods of external evaluation.

Defining the Diploma Examinations

To appreciate the distinctiveness of the diploma examination, it is important to understand not only what the exam is, but also what it is not. The two sections that follow address those definitional concerns.

What the exams are. In the introduction to the 1989-90 Annual Report on the Diploma Examinations Program (Alberta Education, 1990), the stated purposes of the examinations are: (a) to maintain standards of performance, (b) to certify the level of academic achievement of individual students, and (c) to provide province-wide measures of students' achievement in the interest of maintaining student equity.

The diploma examinations are approximately normally distributed with a small negative skew. Many test items are constructed and tested so that they will produce the desired results. Some of the test items are of the multiple choice variety, but there are also some written components. It is erroneous to believe that the examinations identify an objective mastery level that marks students should attain. In fact, if students were generally attaining that level, the examinations would be made more difficult. Unlike a mastery test that separates students into two categories (those who have mastered the topic and those who have not), the diploma examinations are intended to identify the full range of differences in achievement between students. Students are categorized as achieving a

standard of excellence if they score over 80%, an acceptable standard if they score between 50% and 80%, or an implied unacceptable standard (not actually named) if they score below 50%. The diploma examinations are designed to have a mean of about 65%. It is likely the case that if students were generally scoring much higher or lower than that, the examinations would be revised to bring the average back toward 65%. It is this component that makes it impossible to compare students taking the examinations at different sessions and this also undermines the examination's ability to answer whether or not the achievement of students is rising or falling.

Comparisons can be made between students taking the same version of the diploma examination test in the same subject at the same time. There is no mechanism within the examinations to enable comparisons between students who take different examination forms.

One thing Alberta diploma examinations do not overtly claim is the ability to compare students for college or university admission. This is not the case in British Columbia though. Jack Heinrich (1983), Minister of Education in British Columbia, gave three reasons for that province's decision to resume provincial examinations after a 10-year absence: (a) to provide challenging province-wide standards, (b) to ensure a level of equality for graduates applying to colleges and universities, (c) to reassure

the public about the quality of schools. The purposes for reinstatement in British Columbia were nearly the same as in Alberta but with the very notable exception regarding comparison of students for post-secondary studies. In a study on the impact of a return to provincial examinations in British Columbia, one of the recommendations was to "clearly articulate the purposes and the appropriate uses of the provincial examination program" (Anderson et al., 1990, p. 85).

What the exams are not. In Alberta, although the diploma examinations are not directly intended to compare students for university, they indirectly serve that function because they comprise 50% of the final grade for the core subjects. In fact, for some students who take a more academic course load, diploma examinations have more of an impact because more of their courses are diploma examination courses and each of these courses have half of their weight determined by the examinations. Diploma examinations are also not intended as a sort of college admissions or aptitude test, but in practice they are used for that purpose.

Nor are diploma examinations an achievement test of the kind that estimates a student's ability as a percentile rank of some previously measured population. Students taking the same diploma examination at the same time form the population. The percent scores that students receive correspond to the level of achievement on the examination but

factors have been built into the tests which cause them to have a mean of approximately 65% with a standard deviation of between 10-15%.

Although diploma examinations occur at the conclusion of a high school subject, they are not mastery tests. They are based on the curriculum but they are not constructed with the expectation that most students will achieve a mastery level score to demonstrate that they completely understand the curriculum which they have been taught. It is considered that 50% is acceptable on the diploma examinations rather than the 80% or so score that one might expect from a mastery test.

At the other end of the spectrum, diploma examinations are not minimal competency tests. In spite of remarks made that the diploma examinations can be used to ensure a minimum standard, diploma examinations do not do this. Anderson (1983) quotes an Alberta Education policy statement as saying, "a pass mark of 50% is in accord with the public perception of what constitutes a minimum acceptable standard of achievement" (p. 16). This is not to be confused with minimal competency in the American context where the term refers to a basic level of reading and writing that might be termed literacy. Minimal competency tests are often in a form like Florida's functional literacy test (Haney, 1981). A minimal competency test is one in which most students are expected to

pass and demonstrate an adequate level of functioning so that they may graduate from high school.

Although several provinces have a diploma examination program, they are not part of a national testing program. Some countries have a plan of national testing in place that forms the basis for selecting students for universities. Hargadon (1981) cites Butterfield who identifies some of the different national examinations such as the Baccalaureat in France, the Abitur in Germany, and the "A" level examinations in England. The most formidable national college test he identifies is one in China which is "spread over three days, requiring 12.5 hours to complete, and covering six subjects--an examination that is "passed" by only 300,000 of the three million students who take it" (p. 1112-13).

This look at the historical issues, political issues, and definitional issued brings one to the conclusion that diploma examinations are an interesting testing phenomena. They are likely to remain in place for political rather than educational reasons, but they will have educational implications particularly for students planning to apply to colleges and universities.

Perspectives on the Prediction of College Grades

In the United States, predicting which students will be successful in colleges and universities has received considerable study. As far as I can determine, there is no equivalent to the diploma examinations in the United States. In lieu of this, factors such as high school grades, high school rank, achievement test scores, and aptitude test results have been used to improve the quality of selection procedures.

Suggested Predictors

Research opinion tends to be mixed concerning how one can best predict academic success at the college level. One the one hand, there are those like Mouw and Khanna (1993) who conclude that the ability of any predictors to predict college success is disappointingly low. However, the majority of research tends to propose a variety of predictors and combinations of predictors that work reasonably well (Chissom & Lanier, 1975; Halpin, Halpin, & Schaer, 1981; White, Nylin, & Esser, 1985). Other research tends to provide explanations for why some predictors seem to have weaker correlations than expected (Baird, 1984; Weitzman, 1982).

Halpin, Halpin, and Schaer (1981) found that prediction of freshman GPA was best using high school GPA. Following this, achievement tests (California Achievement Tests and ACT Assessment tests) and the

Scholastic Aptitude Test(SAT) were about even in predicting success. In addition, the SAT added to the prediction from high school GPA alone.

However, other studies have have tried different predictors and found different results. Chissom and Lanier (1975) found that significant prediction of freshman GPA could be made from high school grades, SAT-verbal, and SAT-mathematical. By considering two parts of the SAT, each was found to be significant. White, Nylin, and Esser (1985) concluded that the best predictors of college success were number of academic courses in high school, high school rank, high school GPA, and SAT in that order.

These American perspectives often include the SAT in combination with other predictors, the most common of which is high school GPA. Birch (1983) proposed that, in the Canadian context, rather than diploma examinations, he would "opt for either a test of the more general abilities associated with university success or a test based on an agreed upon core (say 25 per cent) of the curriculum in a Grade 12 subject" (p. 70). It is possible that by a general abilities test Birch had in mind something like the SAT. A single test based on a core of the curriculum might be more like an achievement test. Both SATs and achievement tests have been researched in the United States.

In the paragraphs which follow, four of the most frequently proposed predictors from the American literature will provide the focus for the

discussion. These predictors are high school grade, high school rank, the Scholastic Aptitude Test, and achievement tests.

High school grades. High school grades are usually considered the best predictor of college performance. Usually, the variable high school grades involves some average of those grades such as GPA. Some studies have considered both a GPA and individual course grades as predictors.

Some research suggests that high school GPA works much better at predicting college performance than the SAT (Larson & Scontrino, 1976). Slack and Porter (1980) found that the SAT added little to the high school record in predicting college performance. Other research suggests that the combination of high school grades and SATs together are a good predictor of college success (Kaplan, 1982).

Why does some research report a benefit from adding the SAT to high school grades while other research does not? The answer to this question might find a response from Birch (1983). He stated, "in 827 studies, high school grades, when correlated with subsequent university grades, yielded a median correlation coefficient of 0.52; SAT scores, 0.41; and both combined, 0.58." Weitzman (1982) found that after correcting for restriction of range, the SAT may be a highly valid instrument of selection;

in combination with high school record, the corrected coefficient was estimated to be 0.76. So, it would seem that one reason some research found little or no benefit by adding to high school GPA while other research did is that restriction of range can diminish the value of the correlation to the point where it disappears. In fact, Weitzman's correlation of 0.76 is a good deal higher than that suggested by Birch who said that the average was 0.58.

Baird (1984) provides another explanation why some of the research finds little value in adding other variables to high school GPA. He suggested that the strengths of the predictive validities of high school grades and SATs have a great deal to do with the institutions involved For state colleges where students are drawn from a single region and differences between the students in academic performance among the students are greater, the predictions that can be made from the SAT scores are greater. For those colleges that draw a more elite group of students from across the nation, differences among the students are relatively smaller and the ability to make predictions from the SAT is accordingly diminished.

High school rank. High school rank is a variable that rates students from highest to lowest based on their high school grades. In a sense, high school rank is a very abbreviated form of GPA with those with a higher

GPA ranked above those with a lower GPA. Class rank is determined at the high school level.

One finds that the literature involving high school rank is also divided concerning its benefit in combination with other predictors. Baron and Norman (1992) reported that high school class rank and achievement tests added significantly to the prediction of college grades whereas the SAT did not. On the other hand, Hanford (1985) produced results which demonstrated convincingly that class rank by percentile and SAT total scores were quite related and one would therefore suppose that there is merit in each method since they are related to one another.

While studying results of Black female college students in the United States, Houston (1980, 1983) came to two different conclusions. In 1980, Houston wrote that high school rank and SAT scores were not enough to make good decisions for some students. But in 1983, Houston reported that high school rank was more predictive of academic performance than the SAT-verbal or SAT-mathematical.

In the Canadian context, Birch (1983) reported that although the correlation of B.C. high school marks to university grades for the whole province was fairly high (0.52), correlations of high school marks to university grades for limited geographical regions were even higher (0.56 to 0.80). These higher regional correlations imply that ranking is effective

because in a region population grades are more likely comparing student performance to other students in the same local group. He continued to say that rankings across regions was more problematic and that a systematic way of adjusting different regions would be helpful in improving university admissions decisions.

A difficulty with the high school rank issue is that it seems that neither Alberta nor B.C. records high school rank on a student's high school transcript. In order to make use of this variable, data would have to be gathered that are not available even to students themselves.

Scholastic Aptitude Test. The Scholastic Aptitude Test was created to help provide a more uniform basis for admission to colleges in the eastern United States (Schudson, 1972). Although the SAT has been mentioned in regard to high school GPA and high school rank, some research identifies it as an important factor in itself. Mauger and Kolmodin (1975) found that SATs could be used to predict achievement over the entire period of college attendance.

In other research, the verbal or mathematical portions of the SAT were emphasized. Payne, Rapley, and Wells (1973) explained that the chief contributor to the estimation of college academic achievement was the SAT-mathematical. Chissom and Lanier (1975) and Houston (1983)

reported that the verbal and mathematical scores were separate, useful predictive variables.

The SAT is sometimes combined with other tests rather than student grades; in these cases, the SAT usually does quite well. For example Dreher and Singer (1985) found that learning from text, background knowledge, and attitude toward school are factors that can improve prediction above that of the SAT alone.

Criticisms of the SAT had been summarized by Slack and Porter (1980). Among the criticisms were errors in measurement, bias in test items, and secrecy in operation. However, the criticism that SAT scores can be substantially raised by practice and coaching is probably the most serious charge. Slack and Porter (1980) reported that gains of as much as 85 points on the verbal portion and 75 points on the mathematical portion have been demonstrated.

Crouse (1985) stated "my argument is only that most colleges' use of their applicants' SAT score reports does not now help them improve the accuracy of their admissions decisions in terms of academic performance and graduations rates of the students they admit" (p. 219). Crouse and Trusheim (1991) added that the SAT adds minimally in the prediction of college grades. The exceptional cases where the SAT improved the

prediction of student grades seemed to depend on the high school where the student graduated.

Achievement Tests. Achievement tests were developed around the turn of the century. In time correlational methods and mathematics were applied to creating achievement tests which would rank individuals from best to worst (Levine, 1976). Less seems to be written about the use of various achievement tests in predicting college success than is written about the SAT. Perhaps this is because the SAT has a wide usage whereas the achievement tests might be more regional or college-based

In spite of their less frequent use in research, Michael and Shaffer (1979) reported that short tests of English language skills were a better predictor of GPA than the SAT. Baron and Norman (1992) likewise reported that high school class rank and achievement tests added significantly to prediction of college GPA beyond the prediction from high school grades alone whereas the SAT scores did not.

Summary

To summare, high school marks are generally held to be the best predictor of college grades. The amount that SATs and achievement tests improve upon on high school grades varies. The best case can be made for the use of additional predictors when there is greater diversity in the

academic skills of the students being compared such as at a state or regional college.

Prediction of University Grades Using Diploma Examination Marks and Teacher-Assigned Marks

Based on the literature so far, it has been shown that high school grades are probably the most effective predictor of college success. The effectiveness of diploma examinations would be demonstrated to the degree to which they can add to the prediction that could be made from teacher marks alone. In the discussion that follows, the effectiveness of teacher marks and diploma examination marks will be considered followed by the limitations of the diploma examinations marks and the issues involved in combining the two.

Effectiveness of Teacher-Assigned Marks

Birch (1983) summarizes the situation of college grade prediction from teacher marks this way: "More than 800 studies throughout North America are consistent in finding that Grade 12 grades are the best available predictor of university achievement for students entering directly from high school" (p. 69).

Other Canadian research suggests that teacher-assigned marks were more highly correlated with university GPA than diploma exams

(Mehra, 1977). Based on Mehra's work, MACOSA (1979) reported, "teacher-assigned marks provide as good or better predictions of later academic success in post-secondary studies than do marks from common examinations" (p. 12). Birch (1983), when commenting on a study at the University of Alberta (no title given although it could be Mehra's study), stated that it was found after the discontinuation of departmental examinations that

> With the introduction of teacher assessed grading procedure at the high school level, similar to the one employed at the University, the correlation between high school and University grades has inevitably shown improvement. In other words, teachers' grades at the high school level would appear to be somewhat better predictors of grades at the University (p. 70).

In a 1983 study entitled Predictors of Academic Success at Simon Fraser University (reported in Birch, 1983) it was found that the combination of English 12 plus three other subjects was a better predictor than (a) fewer subjects and (b) an average based on all Grade 12 subjects. Grade 12 grades became marginally better predictors after the disappearance of provincial examinations in B.C. (Birch, 1983).

It is unclear if these findings would be true for the Faculty of Education or for the current system where diploma examinations and teacher-assigned grades are each weighted at 50% each.

Effectiveness of Diploma Examination Scores

Unlike the SAT, for which much has been written about how colleges and universities might best avail themselves of information from student scores, the diploma exam is not designed as an admissions test. Nor are the diploma exams like minimal competency tests in that they are not designed to establish or enforce some base level of student skill. While diploma exams are a standardized test and can be used to compare students who take the exam at the same administration, they lack carry over questions that would allow students to be compared who are examined in different years and semesters.

MacDonald (1970) reported that student averages based on departmental examinations correlated moderately with students' university grades. However, the best predictor of any year of university after the first year was the average of grades for the preceding year at university. MacDonald conducted his study with grades from six Atlantic Canadian universities and prior to the availability of separate teacher marks and departmental exam marks. MacDonald also reported that predictions of GPA from departmental examinations varied between faculties. From high school grades, success in the Faculty of Science was found to be most predictable followed in order by the Faculties of Engineering, Arts, and Education (MacDonald, 1970). Of the six Atlantic region universities that

MacDonald considered, correlations between high school grades and university grades differed in the various faculties at the different universities. No correlations were given in his research that aggregate the different university results together.

Diploma Examination Limitations

Alberta Education (1990) describes the purpose of the diploma exams as "to maintain standards of performance, certify the level of academic achievement of individual students and, in the interest of maintaining student equity, provide province-wide measures of students' achievement. The program is designed to support evaluation conducted by teachers in the classroom and, when all factors are considered, the collective results can serve as important information in program monitoring. The diploma examination results indicate achievement of core objectives in accordance with provincial standards. It is limited to those objectives that can be evaluated in a limited time with a paper and pencil test" (p. 1).

While diploma examinations do provide a means to compare students tested on the same exam administration, they are limited by the same factors that limit any paper and pencil test. The teacher-assigned mark, on the other hand, reflects test results and samples of student work over a period of time and on a variety of skills not readily assessed by a

standardized written exam such as oral presentations, group cooperation, and report writing.

Remarking about the discontinuation of departmental examinations in B.C., Birch (1983) commented that "past experience suggests that having a provincial exam count for 100 per cent of the mark... did not provide us with better information on which to predict success" (p. 70). At least in part, departmental examinations were considered in an effort to help deal with the perceived problem of grade inflation (MACOSA, 1977). When teachers were permitted to assign students' grades, there were gradual increases in student averages. On the one hand, this may be viewed as an advantage of diploma examinations because they help to promote fairness and justice in the process of student evaluation (MACOSA, 1977). On the other hand, students presenting marks to the university are not equally affected by the diploma examinations because they do not take the same number of core courses.

Combining Teacher-Assigned Marks and Diploma Examination Marks

Prior to 1973, departmental examinations counted for one hundred percent of a student's grade in that subject. In the period between 1973 and 1983, teacher evaluations of students counted for one hundred percent of the student's grade. When the diploma examinations became mandatory again in 1983, the student's final grade was an average of the

teacher's evaluation and the diploma examination score. This is effectively a weighting of fifty percent for each component. The literature does not suggest an educational reason for selecting a 50-50 weighting.

Researchers adopted a cautious approach to the reinstatement of government examinations. Birch (1983) stated, "we can be fairly certain that making an unpiloted examination count for 50 per cent of the final mark will not enhance the quality of admissions decisions either. If, however, the admissions data on each student included both school grades and marks on a carefully constructed provincial examinations, reflecting the expected curriculum, prediction of university success would surely be enhanced" (p. 70).

Synthesis

In summary, this chapter has considered research that can be arranged in three categories. First, some information is not disputed as a matter of fact. Second, some questions and issues have been raised for which the answers are unknown. Third, some claims have been made that bear further investigation.

Regarding those things which are generally acknowledged, the most foundational is that high school grades are considered the best predictor of college achievement. For the most part, these high school

grades are grades assigned by teachers. In Alberta, the diploma examination grades compose half of the high school grade in core subjects. Alberta's diploma examinations do not claim to be for the purpose of making decisions for university admissions. Diploma examinations are carefully constructed external evaluations but they have the limitations inherent to any paper and pencil test administered in a limited amount of time. These limitations include the narrowing of skills that can be reasonably tested on paper. Diploma examinations are not a minimal competency examination but they exist to certify that the student meets standards that are in keeping with public expectation. When surveyed, students preferred a weighting that would make their teacher-assigned mark worth more than the diploma examination mark. When surveyed, a majority of teachers were of the opinion that the diploma examinations existed more for political rather educational reasons.

It is not known whether teacher marks or diploma examination marks are more effective evaluations of student achievement or are better predictors of subsequent student success in college. It is not known how great a role restriction of range plays in reducing the correlations between university GPA and teacher marks and university GPA and diploma examination marks.

It is claimed that diploma examinations exist for political rather than educational reasons. As diploma examinations are a kind of external evaluation, it is claimed that the improvement in predicting university GPA over teacher marks will depend on the variation in the students. It is expected that for students in a state college or regional university, an external examination would probably produce stronger correlations than for an institution whose students were less diverse. Further, it is claimed that as the regional designations represent smaller constituencies (especially as they would approach something as small as high school class rank) the correlations will be stronger between university GPA and high school mark than the correlations would be over a greater regional area. It is proposed that the combination of teacher marks with diploma examination marks should yield a better correlation with university GPA than either one alone. Short tests of English were found to have some merit in predicting GPA and English in combination with three other subjects was found to be about as good a predictor as the average of all high school subjects were as a predictor.

Research Questions

In an effort to answer some of the unknowns and investigate some of the claims regarding the prediction of college grades from teacher-

assigned marks and diploma examination marks, a number of questions were addressed in this study. The questions were:

1. How well do teacher-assigned marks predict university GPA?
2. How well do diploma examination marks predict university GPA?
3. How similar are teacher marks to diploma examination marks in predicting student achievement?
4. Can these two predictors be combined to better predict university GPA?
5. What is the best combination of teacher marks and diploma examination marks if there is a basis to combine them?
6. If there are differences in the quality of prediction that can be made from different high school subjects, which subject or combination of subjects makes the best prediction?

CHAPTER 3

METHOD

In order to investigate the relative predictive validities of teacher marks and diploma examination marks in predicting university GPA, it was necessary to obtain high school teacher marks and diploma examination marks as well as university GPAs for a large number of students. A student's high school transcript provides both the teacher mark and the diploma examination mark for those courses that have diploma examinations. Until recently, the University of Alberta recorded the diploma examination mark, teacher mark, and final mark for students who were accepted into the university. Therefore, records from the Registrar's Office could provide an excellent source of data to investigate the predictive validities of teacher-assigned marks and diploma examination marks.

Because issues of external evaluation and teacher professionalism are especially relevant to the Faculty of Education and because that faculty is sufficiently large so that there would be many student records, the Faculty of Education seemed a good choice to provide data for this inquiry. As well, the variety of tasks and methods of student evaluation in

the Faculty of Education is likely more similar to the range of evaluation methods that might be reflected in teacher-assigned marks.

Permission was requested and obtained from the General Faculties Council Executive Committee to collect student information for research purposes. Student records were obtained for all students in the Faculty of Education for the 1993-94 school year. To protect the privacy of students, no information was provided that would identify any student.

Sample

A total of 3,679 student records were obtained. Of this number, 2,504 were records for females (68.1%) and 1,175 were records for males (31.9%). Information concerning the program of study revealed that 1,747 (47.5%) of the students were in elementary programs while 1,808 (49.1%) of the students were enrolled in the secondary program. Students in other designations (vocational, industrial, and adult education) accounted for the remaining 3.4% (124 students). All elementary students have the same major. In contrast, secondary students are admitted to particular majors and meet different admissions criteria. The numbers of students in different secondary majors are as follows: Art (14), Biological Sciences (76), Business (32), Drama (20), English (168), General Sciences (75), Home Economics (26), Industrial Education (15), Mathematics (116),

Modern Languages (149), Music (74), Physical Education (271), Physical Sciences (103), and Social Studies (669).

Some students begin their studies in the Faculty of Education in year one, whereas others transfer to the university from colleges that usually offer first and second year courses only, or from other university programs. Therefore there are more students in the later years than in the earlier years of their programs. Based on the highest year with a GPA recorded, there were 1,345 fourth year students, 1,168 third year students, 709 second year students, and 422 first year students. No GPA was available for the remaining 35 students.

Variables

A data matrix was created that included the variables diploma examination and teacher-assigned high school marks in various subjects and university GPA for each year completed in the Faculty of Education at the University of Alberta.

Diploma Examination Marks

Although diploma examination results were available for eight subjects, only six subjects were included because these subjects provided a satisfactory sample size for the intended data analyses. The diploma examination subjects included were English 30, Social Studies 30, Mathematics 30, Biology 30, Chemistry 30, and Physics 30. Neither

English 33 nor Francais 30 was included due to the small number of students that have scores in these courses. The designation "30" is used in the province of Alberta to identify courses that are normally taken in grade 12. Student grades are reported in Alberta as a percentage. These percentage values were entered into the matrix.

Alberta diploma examinations are offered several times each year in the various subjects. Students examined at the same sitting write the same exam for a subject. However, other examinations in the same year and in different years are not the same. There is an intent to make the examinations administered in January and June parallel by using the same Table of Specifications to construct each examination. All reported student grades for a diploma examination subject have been treated as comparable to one another in this research although they are not from the same sitting or from the same year.

Teacher-Assigned Marks

In the province of Alberta, teachers report student marks as a percentage. These values were entered into the data matrix for the same six high school subjects for which diploma examination subjects were included.

As with the diploma examinations, the teachers assigning marks were not the same for all students. However, the teachers were following

the same curriculum and presumably used similar standards for student evaluation. Further, all were familiar with the Table of Specifications for the diploma examinations in their area.

GPA

The University of Alberta reports the student GPA on a 9-point scale. For each year, the GPA is calculated based on the achievement of the student and the weight of the course involved. The values used for GPA were those taken from university records. Students in the same program year do not necessarily have the same courses, same instructors, or the same number of credits over which the GPA was calculated.

Because students can transfer into the Faculty of Education, not every student in each year has GPAs for all the preceding years. The GPAs reported in the data matrix were obtained by students while enrolled in the Faculty of Education at the University of Alberta. By combining the different students from the first, second, third, and fourth year cohorts, there were 1,642 students with a year-one university GPA, 2,257 students with a year-two university GPA, 2,289 students with a year-three university GPA, and 1,345 students with a year-four university GPA.

Sampling Procedures and Sample Size for Various Analyses

A great number of correlation coefficients, multiple regression equations, and other calculations were computed from the available data.

As much as possible, sampling decisions were made to facilitate comparisons between correlations while maximizing the sample size.

Sample Limited to Students with Both a Teacher Mark and a Diploma Examination Mark

Because it is possible for students to have a teacher mark without a diploma examination mark and vice versa, only those student records with both a teacher mark and diploma examination mark were used. This would ensure that comparisons would be made on the same students. It also dramatically reduced the varying number of ns that would appear in many of the tables.

Sample Limited to Students with Year One GPA

In order to determine how well university GPA could be predicted by high school marks, correlations were calculated for GPA in year one with teacher-assigned marks and diploma examination marks for each of the high school subjects. From each subset, the sample for these correlations included all those students for which there was a teacher mark, a diploma examination mark, and a year one GPA mark. This allowed fair comparisons between the correlation of GPA with teacher marks and the correlation of GPA with diploma examination marks, because for each subset the same students were used for both correlations.

The decision was made to aggregate the four different cohorts of students so that the size of n would be as large as possible. Since the teacher-assigned marks, the diploma examination marks, and the GPAs of each cohort all have some variation in evaluation internal to themselves, I felt that this aggregation did not introduce any important, additional complications.

Sampling to Compare Prediction of GPA in Different Years

Because it was likely that the best predictor of GPA in each year of university was the GPA for the year preceding it, correlations were calculated between the GPAs for each of the four years of undergraduate grades. Varying ns were used because selecting only students with records for all four years of GPA would dramatically reduce the number of students included in these calculations.

Once the relationships between the GPAs for different years were established, correlations of teacher marks and diploma examination marks with GPAs in the different years were calculated. These correlations tested whether the predictive validities of teacher marks and diploma examination marks would diminish over the years and whether the predictive validities would diminish similarly. For these correlations, the ns were held constant for teacher marks and diploma examination marks in

the same high school subject and GPA of university year, but they were not held constant across university years or across high school subjects.

Sampling Limitation in the Creation of A Priori Models

For calculating correlation coefficients for the different models that proposed combining teacher-assigned marks and diploma examination marks, new variables were created. For example, for the model that proposed 75% from the teacher mark and 25% from the diploma examination mark, the new composite variable would be 0.75 times the teacher mark for a student plus 0.25 times the diploma examination mark for the same student. For the analyses using these composite variables, the sample was limited to those students who had both a teacher-assigned mark and a diploma examination mark for the same subject; students lacking either the teacher mark or diploma examination mark were excluded from the analyses.

Sampling for Cross-Validation

To cross-validate the empirically derived regression models, the data set was divided approximately in half by a random assigning feature in SPSS. This produced different ns for the different high school subjects because not all students have scores for all subjects.

Sampling Limitations Due to High School Subject

To determine whether there were differences in the predictions that could be made from the six different high school courses, it was necessary to first consider the degree to which the various subjects were correlated with one another. Grade marks for these subjects were generated using the 50/50 model (50% from the teacher mark and 50% from the diploma examination mark). Then correlations were calculated between the grades for the various high school subjects. The ns in these calculations vary because most students do not take all six of these subjects. Rather than reduce the number of students, I decided to calculate the correlations with different ns.

Sampling Limitations When Courses Were Combined

When a single course was used to account for the variance in GPA, the sample size was limited to all those students with a teacher mark, a diploma examination mark, and a GPA for that course. This produced relatively large sample sizes. For example, English alone had an n of 954.

However, when two courses were used to predict GPA, the grades in the predictor courses were averaged; this decrease the sample size to those students who had taken both of those courses. When three courses were used, the three high school grades were summed and divided by three, and the sample was limited to those students who had grades for all

three courses. When all six courses were averaged, the size of n dropped down to 164. Some experience with the data set demonstrated that a method of averaging worked nearly as well as regression. It was decided that the simplicity of averaging had merit both as a technique and because it would be easier to advocate to those who might wish to apply this research in an admissions context.

Because it was noted that the decreasing size of n with more predictor variable account for more of the variance just in itself, a number of the mo re-tested using only the 164 students that had scores for all six s s and GPA in year one.

CHAPTER 4

RESULTS

In this chapter results that answer two broad questions will be reported. The first section addresses the question "How well do teacher marks and diploma examination marks predict university GPA?"; the second section addresses the question "How well do combinations of the grades from high school courses predict university GPA?"

How Well Do Teacher Marks and Diploma Examination Marks Predict University GPA?

One properly expects teacher marks and diploma examination marks to each have some value in predicting future GPA. This section presents results that address how well each type of high school mark predicts university GPA. Predictions of GPA can be arrived at through three different arrangements of teacher marks and diploma examination marks. These arrangements include predictors in isolation, a combination of predictors in a priori models, and a combination of predictors derived from multiple regression equations.

Predictive Strengths of Teacher Marks and Diploma Examinations Marks in Isolation

This section focuses on results that indicate how well teacher marks and diploma examination marks, as isolated variables, predict GPA. First, teacher marks will be evaluated as a predictor of GPA, then diploma examination marks will be considered, and finally, results will be given that compare the predictive validity of high school marks across the four years of university.

Teacher marks alone. If only teacher marks were used to predict GPA, how good would they be? Table 1 shows the number of students in the full sample, the means of and standard deviations for the measures of high school grades, and the correlations between the high school grades in the selected subjects and year one GPA.

Teacher marks correlate with first year GPA from a low of 0.23 (Math 30) to a high of 0.40 (Physics 30). The subject area means, which are very similar, are all in the low or mid-seventies. The subject area standard deviations differ minimally. One might suppose this difference could account for some of the difference in the correlations with GPA, but there is no relationship between the order of the standard deviations and the order of the correlations with first year GPA.

Table 1

Comparison of the Correlations of Teacher Marks and Diploma Examination Marks with GPA in Year One

Course	n	Mean[a]	SD	Correlation[b]	
English	954	6.1	0.91	0.54	
Teacher Mark		75.0	8.15		0 36
Diploma Mark		72.6	8.64		0.38
Social Studies	922	6.1	0.92	0.64	
Teacher Mark		75.3	8.40		0.35
Diploma Mark		71.7	10.10		0.35
Math	828	6.1	0.87	0 66	
Teacher Mark		72.4	11.30		0.23
Diploma Mark		70.2	14.12		0.24
Biology	776	6.1	0.91	0.71	
Teacher Mark		75.4	8.71		0.35
Diploma Mark		74.0	11.69		0.34
Chemistry	738	6.1	0.90	0.66	
Teacher Mark		73.2	9.58		0.29
Diploma Mark		70.9	11.47		0.28
Physics	317	6.2	0.98	0.63	
Teacher Mark		73.2	9.73		0.40
Diploma Mark		68.3	13.23		0.34

[a]The mean and standard deviation on the first line of each group are for the first year GPA. [b]The first correlation in each section is the correlation between teacher mark and diploma examination mark. The second and third correlations are with GPA in year one.

On the whole, teacher marks are only moderate predictors of first year GPA. They would likely be even better predictors if more high school students continued on to university. The students who continue on to university generally have higher and less variable teacher marks than all high school students; this restriction of range will attenuate the correlations.

Diploma examination marks alone. If only diploma examination marks were used to predict GPA, how good would they be? Diploma examination marks correlate moderately with first year GPA. As can be seen in Table 1, correlations ranged from a low of 0.24 (Math 30) to a high of 0.38 (English 30).

The means of the high school subjects vary from the high sixties to the low seventies. Standard deviations range from 8.64 to 14.12. It does not appear that the differences in the standard deviations played a large part in the differences in correlations with GPA, because English 30 has the lowest standard deviation and the highest correlation, whereas Math 30 has the highest standard deviation and the lowest correlation.

On the whole, diploma examination marks are moderate predictors of first year GPA. They also would likely be even better predictors if it were not for attenuation due to restriction of range.

Comparison of the two predictors of GPA. Since teacher marks and diploma examination marks are both reasonably good predictors of first year GPA, comparisons between their relative strengths may be considered. As can be observed from Table 1, for each of the subjects, the mean of teacher grades, although very similar to the diploma exam mean grades, are consistently slightly higher. However, the diploma examinations consistently yield greater standard deviations. Because there are larger standard deviations for diploma examinations, it might be expected that the that diploma examinations would consistentiy produce higher correlations. This is not the case; teacher marks and diploma examination marks provide almost identical correlations for each of the subject areas with GPA. It is possible that if teacher-assigned marks were as differentiated as the diploma examination marks that the correlations of the teacher-assigned marks with GPA would be higher than they currently are.

Why might teacher marks and diploma examination marks offer such similar correlations with GPA? Firstly, for each subject, the teacher mark and the diploma examination mark are correlated with each other more than either is correlated with GPA. The correlations between teacher mark and diploma examination mark in the various subjects ranges from a low of 0.54 to a high of 0.71. This larger correlation is likely

due to a greater similarity of content between the two and a closer proximity in time than either has with first year GPA.

Secondly, it is quite possible that each high school subject can only predict GPA to a limited degree It could not be expected that if the high school subject was itself p. .tive of GPA to a limited degree, that scores from that subject could predict GPA above that limited level. The sir . . of teacher marks and diploma examination marks might each indicate limitations to predict GPA from a given high school subject.

Thirdly, teacher marks have a strength where diploma examinations have a weakness and conversely diploma examinations have a strength where teacher marks have a weakness. Teacher marks are derived from a number of evaluations aggregated together. The process of aggregation leads to less standard deviation if a number of differing skills are considered and students are not similarly gifted across skills. Although a student might be above the mean in one skill, it is possible for them to be below the mean in another and this diminishes the standard deviation. This gives teacher marks an advantage in that they assess more skills over a greater period of time and should make them more like the score represented by GPA. However, teacher marks produce less differentiation between students as reported by the lower standard deviation. Diploma examination marks have greater standard deviations and should produce

greater correlations but they are limited to a single administration and a more restricted range of tasks.

Considering the similarity of correlations with GPA, it would seem that these two different dynamics balance one another. Firstly, differences in these correlations could depend on the relative differences in the number of divergent skills assessed since more skills might lead to less correlation. The case could be made that skills in English are more divergent than skills in the sciences. Secondly, differences are likely due to the relative subjective versus objective nature of evaluation in the different courses. It can be noted that Math and the sciences are considered more objective and these are the ones that are more correlated between the teacher marks and the diploma examination marks.

One the whole, teacher marks and diploma examination marks closely parallel one another in their prediction of GPA. Even more, teacher marks and diploma examination marks closely parallel each other in the same high school course.

Consistency of the predictors over different years. The results presented so far have been limited to the prediction of first year GPA. This section addresses whether or not this is a reasonable restriction. Table 2 summarizes the data related to several issues regarding university GPA. The mean of GPA rises each year throughout the four years and the

Table 2

Consistency of GPA Over the Four Years of University

				Correlation Between Years		
Year	n	Mean	Std Dev	Two	Three	Four
One	1642	6.09	1.02	0.61	0.58	0.52
				(1176)	(726)	(424)
Two	2257	6.40	0.9[illegible]		0.61	0.56
					(1388)	(803)
Three	2289	6.67	0.86			0.65
						(1121)
Four	1345	6.91	0.81			

standard deviation diminishes each year across the four years. This is to be expected for two reasons. First, the university assigns higher mean grades to higher level courses. Second, it is presumed that less able students drop out and that higher achieving students continue to the later years.

The correlations between GPA in any one year with another year reveals that the best predictor for each year is the year which immediately precedes it. Therefore, year one is the best predictor of year two, year two of year three, and year three of year four. Working backwards, it could be proposed that the best predictors of GPAs in year one would be the high school grades immediately preceding them.

Although it would be expected that high school grades would be most strongly correlated with GPA in year one, Table 3 reveals that this is not always the case. In many cases, for both teacher marks and diploma examination marks there is a slightly higher correlation with GPA in years two and three than in year one. Teacher marks are generally somewhat better predictors than diploma examination marks with this trend increasing for later years. By year four, the teacher marks are consistently better predictors than the diploma examination marks and the differences are statistically significant for some subjects. However, the differences between the predictiveness of the teacher marks and the predictiveness of

Table 3

Comparison of the Correlation of Teacher and Diploma Examination Marks With GPA Over Four Years

	Year							
	One		Two		Three		Four	
Course	n	r	n	r	n	r	n	r
English	954		1555		1416		884	*
Teacher Mark		0.36		0.36		0.37		0.39
Diploma Mark		0.38		0.37		0.36		0.32
Social Studies	922		1500		1347		828	
Teacher Mark		0.35		0.38		0.36		0.30
Diploma Mark		0.35		0.36		0.32		0.25
Math	828		1330		1219		769	**
Teacher Mark		0.23		0.28		0.29		0.26
Diploma Mark		0.24		0.29		0.28		0.18
Biology	776		1277	*	1165		717	**
Teacher Mark		0.35		0.40		0.38		0.32
Diploma Mark		0.34		0.36		0.37		0.23
Chemistry	738		1170		1064		682	
Teacher Mark		0.29		0.36		0.35		0.27
Diploma Mark		0.28		0.33		0.32		0.22
Physics	317		515	*	487		352	
Teacher Mark		0.40		0.37		0.37		0.33
Diploma Mark		0.34		0.29		0.31		0.29

Using the test for dependent samples: *Differences are significant at the 0.05 level. **Differences are significant at the 0.01 level.

the diploma examination marks are not significant for the most part especially in the earlier years.

Because of these similarities in prediction across the four years, in the remaining analyses I will limit the reporting of results to GPA in year one. Students enter year one based on their high school marks. If they were to enter at another year by transferring, it would be most sensible to consider their most recent scores rather than more dated scores.

Predictive Strengths of Teacher Marks and Diploma Examination Marks When Combined

Because teacher marks and diploma examination marks are each moderate predictors and because they are correlated with each other but not entirely redundant, it can be asked "Can they be combined in some way to produce better prediction?" In this section, a priori models will be investigated. These models are a priori in that they impose arbitrary weights to the teacher marks and diploma examination marks. Models that will be considered are: (a) the current model that weights teacher marks and diploma examination marks equally (50% each), (b) a model emphasizing teacher marks (75% teacher mark and 25% diploma examination mark), and (c) a model emphasizing diploma examination marks (25% teacher mark and 75% diploma examination mark).

The current model. Currently a student's final reported high school mark in a diploma examination subject weights the teacher mark and the diploma examination mark equally. (This is provided the student has each of these and here I am only considering students who have both scores.)

Results of combining scores in this way are summarized in the center column of Table 4 under the label 50%TM 50%DM (where TM is teacher mark and DM is diploma examination mark). For most high school subjects, the 50/50 combination produces a correlation with GPA in year one that is higher than either the correlation produced by the teacher mark or diploma examination mark alone, which are shown in the first and last columns of Table 4. Only the correlation for Physics 30 teacher mark is equal to the correlation for the pooled mark.

A model emphasizing teacher marks. Results for the model emphasizing teacher marks are summarized in Table 4 under the column heading 75%TM 25%DM. For every course, the results, though not very different, demonstrate a higher correlation with GPA for this combination than either the teacher marks or diploma examination marks alone.

A model emphasizing diploma examination marks. Results for the model emphasizing diploma examination marks are summarized in Table 4 under the column heading 25%TM 75%DM. These results demonstrate

Table 4

Comparison of the Correlations of Proposed Models to Predict GPA from Teacher Marks and Diploma Examination Marks

		Model[a]				
		100%TM	75%TM	50%TM	25%TM	0%TM
Course	n	0%DM	25%DM	50%DM	75%DM	100%DM
English	954	0.36	0.40	0.42	0.41	0.38
Social Studies	922	0.35	0.38	0.39	0.37	0.35
Math	828	0.23	0.25	0.26	0.25	0.24
Biology	776	0.35	0.37	0.37	0.36	0.34
Chemistry	738	0.27	0.31	0.31	0.30	0.28
Physics	317	0.40	0.41	0.40	0.37	0.34

[a]All of the correlation in this table are simple correlations.

that this pooling model is slightly better than either teacher marks or diploma examination marks alone in each case except Physics.

Comparison of the predictive abilities of the combined models.

What conclusions can be drawn from the results of these pooling models? The general observation is that any of the pooling models produces better results than either teacher marks or diploma examination marks alone (except in the case of Physics 30). The improvement due to pooling is because each of teacher marks and diploma examination marks are predictive, each is positively correlated with GPA, and the predictive component in each is not subsumed in the other.

If one were to choose between the pooling models the two best candidates would be the current (50/50) model and the model emphasizing teacher marks (75/25). It should be noted that in three cases the 50/50 combination yields a higher correlation, in two cases they are tied, and only for Physics 30 does the 75/25 model yield a higher correlation. The 50/50 model is superior in more cases and would be the best choice overall. However, none of the models are significantly different from one another based on a test of independent correlations at the 0.05 level.

Optimizing Prediction from Teacher Marks and Diploma Examination Marks

In the preceding section, models were considered that imposed weights on the scores a priori. In this section, a model is considered that derives weights from the scores directly. In the subsections that follow, I will explain how the values were derived, and how they were used, and I will comment on the results obtained.

An optimal model derived from half of the cases. One way to produce an improved model to better predict GPA is to do multiple regression with the variables. But in order to examine the quality of the results, the cases were divided in half. The data from half of the cases were used to derive the weights that would then be tested on the other half of the cases. The cases were randomly divided into two halves using the SPSS program feature. Then using the first half of the cases, a multiple regression analysis was performed for each high school subject using teacher marks and diploma examination marks as the two predictors. These multiple regression results are summarized in Table 5.

Over the different subject areas the GPA means are similar and so are the standard deviations. There is considerable variation in the constants (the y-intercepts in the regression equations) that result, from a low of 2.94 in English 30 to a high of 4.80 in Math 30. This indicates that

Table 5

The Optimal Model Derived to Predict GPA in Year One from Half of the Cases

Course	n	Mean	SD	Constant	Beta[a]	r^2	R^2
English	487	6.09	0.93	2.94	36/64		0.12
Teacher Mark					0.14	0.07	
Diploma Mark					0.25	0.10	
Social Studies	466	6.06	0.94	3.33	66/34		0.11
Teacher Mark					0.23	0.10	
Diploma Mark					0.12	0.08	
Math	414	6.09	0.87	4.80	17/83		0.08
Teacher Mark					0.05	0.04	
Diploma Mark					0.24	0.08	
Biology	389	6.05	0.92	3.55	49/51		0.11
Teacher Mark					0.17	0.09	
Diploma Mark					0.18	0.10	
Chemistry	378	6.05	0.91	3.79	50/50		0.10
Teacher Mark					0.17	0.08	
Diploma Mark					0.17	0.08	
Physics	153	6.18	0.95	3.52	55/45		0.16
Teacher Mark					0.24	0.14	
Diploma Mark					0.20	0.13	

[a]The first digits in each section represent the fractional components of the beta weights so that the teacher marks and diploma examination marks might be given as percents that sum to 100%.

students with Math likely have higher minimums than students who do not have Math. Their mean grades are not particularly higher though. Another reason why the constant for Math 30 may be higher is that the correlation is lower between Math and GPA and therefore slope of the line is less steep.

The relative difference in the beta weights for teacher mark and diploma examination mark hints at what the ideal weighting of each would be. The beta weights were used to calculate how the teacher marks and diploma examination marks might be combined so that they could be compared to the a priori models. For example, in Social Studies 30, the teacher mark has a beta weight of 0.23 and the diploma examination mark has a beta weight of 0.12. By summing the 0.23 with the 0.12, one gets 0.35. If the 0.23 (teacher mark) and 0.12 (diploma examination mark) are each divided by 0.35 (sum of the two), the result is 0.66 or 66% weight from the teacher marks and 0.34 or 34% from the diploma examination mark. These fractional values range from favoring teacher marks in Social Studies 30 (66/34) to favoring diploma examination marks in Math 30 (17/83). However for the sciences the weights are nearly equal

Also in Table 5, r^2 indicates the amount of variance in GPA that can be accounted for by each of the teacher marks or diploma examination marks alone. There is no consistent trend favoring teacher marks or

diploma examination marks in accounting for more variance in GPA. R^2 in Table 5 represents the amount of variance that can be accounted for by combining the teacher mark and the diploma examination mark. The two r^2 values are not simply added to produce R^2, because much of the variance in GPA accounted for by the teacher mark overlaps with the variance accounted for by the diploma examination mark.

Testing the optimal weights with the cross-validation cases. Using the beta weights and constant derived from multiple regression on the original half of the cases, new predictions of GPA were made for the second half of the cases. Table 6 indicates the amount of variance in actual GPA that was accounted for by the predicted GPA from each half of the cases. The column labeled Cross-Validation in Table 6 summarizes the amount of variance in GPA that was accounted for by the predictions made with the second half of the cases using the prediction model derived from the first half of the cases.

For four subjects (Math 30, Biology 30, Chemistry 30, and Physics 30), the amount of variance accounted for is similar to that accounted for in the original cases. In English and Social Studies, the original model seems to account for a good deal more of the variance in the cross-validation cases, even though the regression values are derived from the original data. An explanation for this result follows in the next section.

Table 6

Comparison of Correlations for the Optimal Weighting of Marks with GPA in Year One

	Samples					
	Original			Cross-Validation		
Course	Mean	SD	R^2	Mean	SD	R^2
English			0.12			0.25
Actual	6.09	0.93	(487)	6.07	0.88	(467)
Predicted	6.09	0.32		6.09	0.32	
Social Studies			0.11			0.20
Actual	6.06	0.94	(466)	6.11	0.87	(456)
Predicted	6.06	0.31		6.11	0.30	
Math			0.08			0.05
Actual	6.09	0.87	(414)	6.11	0.87	(414)
Predicted	6.09	0.24		6.09	0.23	
Biology			0.11			0.16
Actual	6.05	0.92	(389)	6.12	0.91	(387)
Predicted	6.05	0.30		6.06	0.32	
Chemistry			0.10			0.10
Actual	6.05	0.91	(378)	6.06	0.90	(360)
Predicted	6.05	0.28		6.06	0.27	
Physics			0.16			0.17
Actual	6.18	0.95	(153)	6.13	1.00	(164)
Predicted	6.18	0.39		6.11	0.35	

Explanation of inconsistencies in the cross-validation sample. R^2 is the amount of variance in GPA that was accounted for. In cross-validation, the value of R^2 is expected to shrink depending on the size of the ns involved. In samples of this size, the shrinkage would be small.

Why are the R^2 for English 30 and Social Studies 30 higher in the cross-validation sample than in the original sample, even though one would expect them to be similar to the original sample if not slightly smaller? The best explanation is random sampling variation. Even though the whole data set was randomly divided and should have been roughly equal, it turns out that with regard to English and Social Studies the one group was distributed in such a way that it favored a higher correlation. Table 9 which will be discussed later hints at the explanation for this result. When multiple regression is done on the whole set of data, the variance accounted for in English 30 is at 0.18 rather than the low of 0.12 in the original or 0.25 in the cross-validation. It seems that by chance more of the student scores that were predictive ended up in the cross-validation sam

Why did the beta weights derived from the first half of the data do so well predicting the English 30 and Social Studies 30 scores if there was a difference in the two random samples? Table 7 gives a good indication to the main reason. Since the beta weights tended toward the middle

Table 7

Comparison of the Predictiveness of Optimal Weights and 50/50 Weights to Predict GPA in Year One Using All the Cases

		Beta Weights		%Weight	R with GPA[a]	
Course	n	TM	DM	TM/DM	Opt.	50/50
English	954	0.22	0.26	45/55	0.42	0.42
Social Studies	922	0.22	0.21	51/49	0.39	0.39
Math	828	0.12	0.16	42/58	0.26	0.26
Biology	776	0.23	0.17	57/43	0.37	0.37
Chemistry	738	0.17	0.17	51/49	0.31	0.31
Physics	317	0.31	0.14	69/31	0.41	0.40

[a]For each subject, the pair of correlations are each with GPA in year one. The optimal column is the correlation formed from using the weights from teacher marks and diploma exam marks derived from multiple regression (These weights are listed in the %Weight column). The 50/50 column reflects a combination where the teacher mark and diploma exam mark are weighted equally or 50% each.

rather than grossly favoring either teacher marks or diploma examination marks and since the a priori models indicate that better predictions are made by combining, then even these weights which do not best fit the cross-validation sample still produce a higher correlation than either the teacher marks or diploma examination marks alone. Because the variance accounted for is the square of the correlation, if there is a higher correlation, then there will be more variance accounted for.

While the cross-validation was consistent for the sciences, for those subjects it should be noted that the beta weights were approximately equal for the teacher marks and diploma examination marks. If the original sample had yielded beta weights for English and Social Studies that were also about 50/50, then the R^2 for each sample would have been much more similar.

A model derived by multiple regression is only good to the degree to which the sample it is based on is representative of the population to which the weights from the regression will be applied. So this seems to be the caveat with deriving weightings to predict GPA from teacher marks and diploma examination marks. If the sample (even though large and random), does not quite reflect the whole then the prediction will be off. Unfortunately, without cross-validation one can not always know how far off true a model may be.

Deriving optimal weights from all of the data together. In order to construct an optimal composite for each high school course, both halves of the data were used together. From differences between the amount of variance accounted for in original sample and in the cross-validation, one could suppose that by using the whole data set that the under-prediction in the one group might be evened out by the over-prediction in the other group.

Table 7 summarizes the results of deriving weightings from the whole data set. The columns headed by "Beta Weights" lists the comparative weights for the teacher mark (TM) and the diploma exam mark (DM) for each subject so that the best prediction of GPA might be obtained. The %Weight column gives a conversion of the beta weights so that one might compare as a percentage how much each of the teacher mark and diploma exam mark contribute to the optimal combining of the two.

In English, for example, the teacher mark has a beta weight of 0.22 and the diploma exam mark has a beta weight of 0.26. On can tell by the similarity of these values that the optimal weight will be about an equal share of each. To change the beta weights to a percentage, each beta weight was divided by the sum of the two for the same subject. So for the English 30 teacher mark, the value equals 0.22 divided by 0.48 (the sum of

0.22 and 0.26) or 45% while the diploma exam percentage is 0.26 divided by 0.48 or 55%.

For five of the subjects (English 30, Social Studies 30, Math 30, Biology 30, and Chemistry 30), one can readily see that the optimal weighting is very nearly 50/50. Three subjects slightly tend in favor of using more of the teacher mark (Social Studies 30, Biology 30, and Chemistry 30) while two (English 30 and Math 30) tend in favor of using more of the diploma exam mark. The optimal combination for Physics 30, unlike the other subjects, favors the teacher mark by more than twice as much as the diploma exam mark (69% teacher mark to 31% diploma exam mark).

The rightmost column in Table 7 compares the correlation of GPA with the optimized mark and the 50/50 weighted mark. In most cases, the correlations are the same. That means that a 50/50 weighting predicts as well as a weighting derived from doing multiple regression.

How Well Do Combinations of the Grades From High School Courses Predict University GPA?

Although the various high school courses have been mentioned earlier in this chapter, those discussions had the comparison of the predictive validities of teacher marks and diploma exam marks as the

central issue. The focus here is somewhat different in that high school course is considered as a moderator in the prediction of GPA. To compare the predictive validities of the courses, for each student, a single grade was calculated for each high school course. It seemed best to weight the teacher mark and diploma exam mark equally. There are two good reasons for this decision: (a) student grades are normally reported as the average of the teacher mark and the diploma exam mark (a 50/50 weighting) and (b) from Table 7, it was apparent that the 50/50 weighting was as correlated as well with GPA as the weighting derived from multiple regression.

In considering the predictive differences that exist between the courses, I must emphasize that there are probably more similarities than differences. Table 8 summarizes the correlations that exist between the subjects. The lowest correlation between any two subjects is between English 30 and Math 30 (0.36) while the highest correlation is between Chemistry 30 and Physics 30 (0.68). Most of the correlations are about 0.60 with many of these being between the various sciences. So any discussion about the moderating effect of the different high school courses must begin with an acknowledgment that there is much that is similar in student performance in different subjects.

Table 8

Correlations Between High School Subjects When Each Course Has A 50/50 Weighting

	High School Subject					
Subject	English	S.S.	Math	Biology	Chem.	Physics
English		0.65	0.36	0.54	0.43	0.47
	(2014)[a]	(1844)	(1673)	(1580)	(1461)	(681)
Social			0.43	0.63	0.53	0.49
Studies		(1929)	(1619)	(1535)	(1417)	(646)
Math				0.54	0.67	0.67
			(1726)	(1382)	(1372)	(666)
Biology					0.64	0.64
				(1646)	(1194)	(500)
Chemistry						0.68
					(1512)	(592)
Physics						
						(702)

[a]The number in each of the parentheses indicates the number of students involved for the calculation of the correlation. The numbers down the diagonal indicate the number of students that had both a teacher mark and a diploma exam mark for each subject

A good method to compare the courses involves determining how much of the variance in GPA is accounted for using each subject as a predictor. One has to keep in mind that these courses do not primarily exist for the sake of predicting GPA but they are utilized in this way in Table 9

Table 9 presents a number of grouping of courses which when combined improve on the prediction that can be made by one course alone. The column labeled "Course" identifies which of the high school subjects is part of the various models. The first letter of the course name is given (E is for English 30, S for Social Studies 30, and so on). Under the "Correlation" column, the sub-heading "TM" stands for teacher marks, "DM "for diploma examination marks, and "Both" is when the average of the teacher and diploma examination marks is used. "Variance Accounted" heads the columns which contain the percentage of the variance in first year GPA that is accounted for by the predicting variables. Going down Table 9, the first six models present the courses in isolation, the next seven models are combinations of courses, and the last seven models are for comparison when the size of n is held constant.

Generally, the right-most column contains the value that accounts for the most variance. The predictors in this column are both the teacher marks and diploma examination marks averaged together. By making comparisons from this column, one can observe that English 30 actually

accounts for more variance in GPA than any other course. Following English 30 (18%), in order of variance accounted for, are Physics 30 (16%), Social Studies 30 (15%), Biology 30 (14%), Chemistry 30 (10%), and Math 30 (7%). English 30 is significantly more correlated with university GPA than Math 30 ($z=3.36$) and Chemistry 30 ($z=2.59$). Social Studies 30 ($z=2.59$), Biology 30 ($z=2.00$), and Physics 30 ($z=2.05$) are also significantly more correlated with university GPA than Math 30.

It should be noted that the predictions made from teacher marks and diploma examination marks account for quite similar amounts of the variance in first year GPA. The combination of the two together is usually only slightly more than either one separately.

The middle of Table 9 considers how much more variance can be accounted for by adding courses to English 30. English 30 is included in every model in the middle of the table for two reasons. First, almost all students take English 30. Second, since English 30 is the best overall predictor and significantly better than both Chemistry 30 and Math 30, it is the one most worth including.

For example, in the right-most column, adding Social Studies 30 (15%) to English 30 (18%) only produces a result of 21% because much of the variance accounted for by English is overlapped with the variance

Table 9

Comparison of the Variance Accounted for by High School Subjects in the Prediction of GPA in Year One

Model	Course[a] E	S	M	B	C	P	n	Correlation[b] TM	DM	Both	Variance Accounted TM	DM	Both
Model 1	X						954	0.36	0.38	0.42	13%	14%	18%
Model 2		X					922	0.35	0.35	0.39	12%	12%	15%
Model 3			X				828	0 23	0.25	0.28	5%	6%	7%
Model 4				X			776	0.35	0.34	0.37	12%	11%	14%
Model 5					X		738	0.29	0.28	0.31	8%	8%	10%
Model 6						X	317	0.40	0.34	0.40	16%	11%	16%
Model 7	X	X					869	0.42	0.42	0.46	17%	17%	21%
Model 8	X		X				802	0.37	0.40	0.43	14%	16%	18%
Model 9	X	X	X				761	0.43	0.45	0.49	19%	20%	24%
Model 10	X	X	X	X			612	0.46	0.47	0.50	21%	22%	25%
Model 11	X	X	X		X		628	0.47	0.47	0.51	22%	22%	26%
Model 12	X	X	X			X	275	0.53	0.52	0.56	28%	27%	32%
Model 13	X	X	X	X	X	X	164	0.62	0.52	0.61	39%	27%	37%
Model 1a	X						164	0.56	0.39	0.53	32%	15%	28%
Model 2a		X					164	0.51	0.37	0.48	26%	13%	23%
Model 3a			X				164	0.35	0.41	0.42	12%	17%	17%
Model 4a				X			164	0.42	0.28	0.38	17%	8%	14%
Model 5a					X		164	0.49	0.38	0.48	24%	14%	23%
Model 6a						X	164	0.46	0.38	0.47	21%	15%	22%
Model 9a	X	X	X				164	0.59	0.51	0.60	35%	26%	36%

[a]An X in the row indicates that the high school subject is included in the correlation.
[b]All of these correlations are simple correlations.

accounted for by Social Studies 30. This overlap was expected from Table 8 since English 30 and Social Studies 30 are correlated with each other 0.65. Adding Math 30 (7%) to this brings the total to 24%. It would seem that although Math does not account for very much variance in itself, the variance it accounts for is somewhat distinct compared to English 30 and Social Studies 30 together. In other words, because the skills one uses in Social Studies 30 are much like the skills one uses in English 30 and quite different from the skills one uses in Math 30, Math 30 ends up adding to the total.

From these three core subjects (24%), adding science to these three core subjects, can produce increases in the variance explained. Adding Biology 30 (14%) yields 25%. Adding Chemistry 30 (10%) yields 26%. And adding Physics 30 (16%) yields 32%. One might wonder what happens with all six subjects. The result is that 37% of the variance is accounted for.

Well, if all of these subjects together can do such a good job of predicting GPA, why not make predictions that way? The problem is that most students do not have all of these subjects. One can see how the n drops as subjects are added. Even though Physics 30 is quite a good predictor, most students do not have that subject.

It becomes a question of making good predictions with the information available. Because most students do have English 30, Social Studies 30, and Math 30, one has to consider whether the addition of more variables is worthwhile if the model begins to be useful for fewer and fewer students. Of the seven multi-course models given in the middle of the table, Model 9 seems to be a good choice in that it (a) includes the three most frequently taken courses, (b) accounts for 24% of the variance, and (c) is not dramatically improved upon when the drop in the size of n is considered (compare Model 9a with Model 13).

The bottom section of Table 9 contains several models with the designation '0' in the number. These models are comprised only of the 164 students who were eligible to be included in every model. These models are included to allow for observations on the differences between courses when the particular students involved are constant. Even though there is a rise in the variance accounted for by as much as 10%, English 30 and Social Studies 30 continue to be the most important courses in accounting for variance while Math 30 remains one of the poorer predictor in isolation.

Among these 164 students, the model including English 30, Social Studies 30, and Math 30 accounts for 36% of the variance, whereas the model including every course only accounts for 37%. This would seem to

be another strong argument in favor of making decisions from English 30, Social Studies 30, and Math 30. It is quite interesting that among these students that have taken all six of the core subjects, the teacher-assigned mark usually accounts for much more variance in first year GPA than the diploma examination marks does.

CHAPTER 5

CONCLUSIONS

This chapter is divided into four sections. The first section contains a summary of the findings. The second section is concerned with the limitations in this research. This is followed by some implications and suggested applications of the research. The final section outlines recommendations for further research.

Summary of Findings

The findings are summarized and organized under two headings. The first division deals with how well teacher marks and diploma examination marks predict university GPA. The second division considers the combinations of high school courses that can be used to predict university GPA.

How Well Do Teacher Marks and Diploma Examination Marks Predict University GPA?

The marks from teacher-assigned marks in the various high school courses examined correlated between 0.30 to 0.40 with university GPA for first year students. These values are somewhat lower than the literature suggests the correlation between high school grades and college GPA

tends to be (Birch, 1983). The diploma examination marks also correlated with university GPA about 0.30 to 0.40.

Teacher marks seem to be slightly better predictors than diploma examination marks especially if one considers correlations with GPA in the fourth year when some of these small differences become statistically significant. The best predictor of GPA in any one year is the GPA of the preceding year.

The correlations between teacher marks and diploma examination marks in the various high school subjects are higher than either the correlations between teacher marks and university GPA or the correlations between diploma examination marks and university GPA. These correlations between teacher marks and diploma examinations marks range from 0.54 (English 30) to 0.71 (Biology 30).

As might be anticipated from the literature, the combination of teacher marks and diploma examinations marks produced better correlations with GPA than either one alone. Of the a priori models examined, the model that weighted teacher marks and diploma examination marks 50/50 produced the highest correlations with university GPA. However, the differences between the models were small and not significant.

When multiple regression models that combined teacher marks and diploma examination marks were examined, these regressions produced results there were essentially the same as the 50/50 a priori model. For most of the subjects, deriving a prediction from multiple regression did not improve on the prediction that could be made by simply averaging the teacher mark and diploma examination mark.

The most important finding was that teacher-assigned marks and diploma examination marks were quite similar in their predictive validity. Because the best a priori model (50/50) does about as well as the regression model, I would conclude that the 50/50 weighting is best, when various things are considered. This choice of weighting appears to have educational merit in the practical prediction of university GPA. On the other hand, the increase in the degree of prediction over either teacher marks or diploma examination marks alone is small and not statistically significant.

How Well Do Combinations of the Grades From High School Courses Predict University GPA?

When the six core high school subjects were considered separately as predictors of university GPA, English 30 was the best predictor and Math 30 is the poorest. English 30 was significantly more correlated with GPA than either Math 30 or Chemistry 30 were. Math 30 was significantly

less correlated with university GPA than English 30, Social Studies 30, Biology 30, or Physics 30. Because tests of English were identified in the literature as contributing to college grade prediction above what could be predicted by an average of high school grades alone, this result is not surprising. It might be surprising that Math 30 is the poorest predictor Because It is likely that most Education students are not taking many mathematics courses or math related courses at university, that Math 30 is probably less predictive than it might be for students in some other faculties such as science, engineering, or business.

When considering how the core high school subjects could best be combined for predictive purposes, it is important to consider the number of students that present grades for each of the courses. Ordered from greatest number of students to fewest, the courses were English 30, Social Studies 30, Math 30, Biology 30, Chemistry 30, and Physics 30. A model that weighted English 30, Social Studies 30, and Math 30 equally accounted for almost as much of the variance in university GPA as a model that weighted all six core subjects equally. When differences in the number of students presenting various courses were considered, this model appeared to be the most appropriate one.

For most students, an average of teacher marks and diploma examination marks together accounts for more of the variance than either

teacher marks or diploma examination marks alone in the prediction of university GPA. An exception to this was noticed among the students who had taken all six core subjects and in their case, the teacher marks were sometimes correlated more highly with university GPA than the combination of teacher marks and diploma examination marks were correlated with GPA but these differences are not statistically significant.

Limitations

There are several limitations in the research I have presented. The greatest problem is the restriction of range. As the literature suggested, higher correlations could be achieved if it were possible to account for the performance of students who applied to but were not accepted to the university. Obviously, those students who did not come to the Faculty of Education cannot be accounted for and this reduces the correlations because of attenuation.

My results reflect the situation in the Faculty of Education, the findings might be different for students in other faculties and programs of study.

Many factors that have an effect on student GPA were not studied. No attempt was made to consider the role student goals and ambitions play in university achievement. Emotional factors and personality traits

were also not considered. These were excluded not because they are unimportant but because they were simply beyond the scope of this research.

These results do not address the degree to which teachers or diploma examinations are functioning in their primary role as evaluators of student performance for particular high school subjects. Comparisons between teacher-assigned marks and diploma examination marks need to be regarded in the context that this research is about the prediction of university GPA. Neither the teachers nor the examinations are primarily concerned with selecting students for university.

It is not possible to determine how much the institution of the diploma examinations has changed teachers' evaluation practices to make them more consistent with the diploma examinations. Because the examinations have probably had an impact on teachers, one cannot definitively answer whether teacher marks or diploma examination marks would have the same quality of prediction in the absence of the other.

Although diploma examinations do not significantly improve the prediction that can be made from teacher marks alone, they probably have the effect of standardizing the curriculum and the methods of evaluation that were intended. The moderate correlations of the teacher marks with

university GPA might partly result from the existence of the diploma examinations.

Implications and Applications

As a result of my research, I have five recommendations. First, for the sake of future research into the quality of prediction that can be made from teacher marks and diploma examination marks, I recommend that the university record both the student's teacher-assigned mark and diploma examination mark.

Second, in those cases where a student has borderline qualifications for admission, because English is the best of the high school predictors and because most students have a grade for English, I believe that the English mark should receive particular consideration.

Third, the importance of English skills for university success should be emphasized to high school students.

Fourth, it should be emphasized to teachers that their evaluations are as valid predictor as the diploma examinations.

Fifth, it is important to inform the public of the actual quality of prediction that teacher marks and diploma examination marks have. This includes acknowledging that although these marks are comparable

indicators of university success, much of the variance in university achievement is not accounted for by either of these predictors.

Recommendations for Further Research

I have three recommendations for further research. First, similar research could be conducted in other faculties to see if the results are the same. If the results are different, the reasons for the differences would be important to discover.

Second, it would be interesting to determine whether or not the combination of English 30, Social Studies 30, and Math 30 alone would provide a good basis for decision making as the current admissions policy.

Third, it would be useful to determine if a regional ranking system was effective for making predictions of university achievement beyond the predictions that can be made from teacher marks and diploma examination marks. If a regional ranking were not a useful additional factor, this could help to demonstrate that the diploma examination process has standardized the province to the point where regionalism is not a factor.

BIBLIOGRAPHY

Alberta Education. (1990). Annual Report: Diploma examinations program. Edmonton, Alberta: Alberta Education.

Airasian, P.W., Madaus, G.F., & Pedulla, J.J. (1979). Minimal Competency Testing. Englewood Cliffs, N.J.: Educational Technology Publications.

Anderson, J.O., Muir, W., Bateson, D.J., Blackmore, D., & Rogers, T. (1990). The impact of provincial examinations on education in British Columbia: summary report. Victoria, B.C.: British Columbia Department of Education.

Anderson, J.O., Muir, W., Bateson, D.J., Blackmore, D., & Rogers, T. (1990). The impact of provincial examinations on education in British Columbia: general report. Victoria, B.C.: British Columbia Department of Education.

Anderson, S.M. (1983). Student evaluation: The political or the educational approach? The Trustee, 53(2), 14-16.

Baird, L.L. (1984). Predicting predictability: The influence of student and institutional characteristics on the prediction of grades. Research in Higher Education, 21(3), 261-279.

Baron, J., & Norman, M.F. (1992). SATs, achievement tests, and high-school class rank as predictors of college performance. Educational and Psychological Measurement, 52(4), 1047-1055.

Birch, D.R. (1983). A researcher's view: How well do exams predict achievement in university? B.C. Teacher, 63(2), 69-70.

Buckley, C. (1983). What kind of student evaluation do parents want? The Trustee, 53(2), 10-11.

Byfield, T. (1983). Education's acid test. Canadian School Executive, 2(9), 7-8.

Calder, P. (1990a). Impact of diploma examinations on the teaching-learning process. Edmonton, Alberta: Alberta Teachers' Association.

Calder, P. (1990b). Impact of external examinations on the teaching-learning process: A review of the literature. Edmonton, Alberta: Alberta Teachers' Association.

Chissom, B.S., & Lanier, D. (1975). Prediction of first quarter freshman GPA using SAT scores and high school grades. Educational and Psychological Measurement, 35(2), 461-463.

Crouse, J. (1985). Does the SAT help colleges make better selection decisions? Harvard Educational Review, 55, 195-219.

Crouse, J., & Trusheim, D. (1991). How colleges can correctly determine selection benefits from the SAT. Harvard Educational Review, 61, 125-147.

Dreher, M., & Singer, H. (1985). Predicting college success: Learning from text, background knowledge, attitude toward school, and the SAT as predictors. National Reading Conference Yearbook, 34, 362-368.

Ferguson, L. (1983). Old wine in new bottles. The Trustee, 53(2), 29-30.

Halpin, G., Halpin, G., & Schaer, B.B. (1981). Relative effectiveness of the California Achievement Tests in comparison with the ACT Assessment, College Board Scholastic Aptitude Test, and high school grade point average in predicting college grade point average. Educational and Psychological Measurement, 41(3), 821-827.

Haney, W. (1981). Validity, vaudeville, and values: A short history of concerns over standardized testing. American Psychologist, 36, 1021-1034.

Handford, G.H. (1985). Further comment: Yes, the SAT does help colleges. Harvard Educational Review, 55, 324-331.

Hargadon, F. (1981). Tests and college admissions. American Psychologist, 36, 1112-1119.

Hart, C. (1988). The diploma dilemma. The ATA Magazine, March/April, 11-13.

Heck, G. (1983). In quest of improved standards. The Trustee, 53(2), 19-21.

Heinrich, J. (1983). The Ministry of Education's view: Why provincial examinations? B.C. Teacher, 63(2), 59-60.

Houston, L.N. (1980). Predicting academic achievement among specially admitted Black female college students. Educational and Psychological Measurement, 40(4), 1189-1195.

Houston, L.N. (1983). The comparative predictive validities of high school rank, the Ammons Quick Test, and two scholastic aptitude test measures for a sample of Black female college students. Educational and Psychological Measurement, 43(4), 1123-1126.

Huntley, J.F. (1976). Academic evaluation and grading: An analysis and some proposals. Harvard Educational Review, 46, 612-631.

Jamha, A. (1988). Role of the Minister of Education in Alberta. Unpublished master's thesis, University of Alberta, Edmonton, AB.

Kaplan, R.M. (1982). Nader's raid on the testing industry: Is it in the best interest of the consumer? American Psychologist, 37(1), 15-23.

Killough, E.A. (1983). A superintendent's view: Not again! B.C. Teacher, 62(2), 55-58.

Larson, J.R., & Scontrino, M.P. (1976). The consistency of high school grade point average and of the verbal and mathematical portions of the Scholastic Aptitude Test of the College Entrance Examination Board, as predictors of performance: An eight year study. Educational and Psychological Measurement, 36(2), 439-443.

Levine, M. (1976). The academic achievement test -- Its historical context and social functions. American Psychologist, 31, 228-238.

MacDonald, E. (1970). Prediction of university success in years after the first. Unpublished master's thesis, University of Alberta, Edmonton, AB.

Martin, G. (1983). A new era in Alberta school testing. The Trustee, 53(2), 6-9.

Martin, G. (1987). Education in Alberta is passing the test. Edmonton Sunday Sun, June 28, 13.

Martin, R. (1983). Comprehensives: A step in the wrong direction. The Trustee, 53(2), 25-27.

Mauger, P., & Kolmodin, C.A. (1975). Long-term predictive validity of the Scholastic Aptitude Test. Journal of Educational Psychology, 67(6), 847-851.

Mehra, N. (1977). High school grades and university performance. Edmonton, Alberta: Office of Institutional Research and Planning University of Alberta.

Michael, W.B., & Shaffer, P. (1979). A comparison of the validity of the Test of Standard Written English (TSWE) and the California State University and Colleges English Placement Test (CSUC-EPT) in the prediction of grades in a basic English composition course and overall freshman-year grade point average. Educational and Psychological Measurement, 39(1), 131-145.

Minister's Advisory Committee on Student Achievement (MACOSA). (1977). Alberta grade 12 examination study condensed version. Edmonton, Alberta: Department of Education.

Minister's Advisory Committee on Student Achievement (MACOSA). (1979). Student achievement in Alberta study executive summary. Edmonton, Alberta: Department of Education.

Minister's Advisory Committee on Student Achievement (MACOSA). (1980). Public reaction to the recommendations of the Minister's Advisory Committee on Student Achievement. Edmonton, Alberta: Department of Education.

Minister announces new student evaluation policy. (1983). The Trustee, 53(2), 3.

Mouw, J.T., & Khanna, R.K. (1993). Prediction of academic sucess: A review of the literature and some recommendations. College Student Journal, 27(3), 328-336.

Muir, P. (1983). The days of liberalization are over. The Trustee, 53(2), 4-5.

Nemeth, J., & Samiroden, W. (1990). Diploma exams: A political approach to evaluation. ATA Magazine, 70(3), 28-31.

Payne, D.A., Rapley, F.E., & Wells, R.A. (1973). Application of biographical data inventory to estimate college academic achievement. Measurement and Evaluation in Guidance, 6(3), 152-156.

Schudson, M.S. (1972). Organizing the meritocracy: A history of the College Entrance Examination Board. Harvard Educational Review, 42, 34-69.

Slack, W.V., & Porter, D. (1980). The Scholastic Aptitude Test: A critical appraisal. Harvard Educational Review, 50(2), 154-175.

Stuart, R. (1987). The abuse of diploma exams in Alberta. The Alberta Science Teacher, 7(3), 2-6.

Symyrozum, L., & Hrabi, J. (1984). Student evaluation. The Trustee, 54(1), 26-28.

Tomlinson, D. (1983). A step in the wrong direction. The Canadian School Executive, 2(9), 9.

Weitzman, R.A. (1982). The prediction of college achievement by the Scholastic Aptitude Test and the high school record. Journal of Educational Measurement, 19(3), 179-191.

White, W.F, Nylin, W.G., & Esser, P.F. (1985). Academic course grades as better predictors of graduation from a commuter-type college than SAT scores. Psychological Reports, 56(2), 375-378.