PRODUCING EQUIVALENT EXAMINATION FORMS: AN ASSESSMENT OF THE BRITISH COLUMBIA MINISTRY OF EDUCATION EXAMINATION CONSTRUCTION PROCEDURE

by

PETER DENTON MAC MILLAN

B.Sc., University of British Columbia 1972

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ARTS

in

THE FACULTY OF GRADUATE STUDIES

Department of Educational Psychology

We accept this thesis as conforming to the required standard

©THE UNIVERSITY OF BRITISH COLUMBIA

August 1991
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Educational Psychology

The University of British Columbia
Vancouver, Canada

Date Sept. 6, 1991
ABSTRACT

Questions have been raised concerning the equivalency of the January, June, and August forms of the British Columbia provincial Grade 12 examinations for a given subject. The procedure for constructing these examinations has been changed as of the 1990/91 school year. The purpose of this study was to duplicate this new procedure and assess the equivalency of the forms that resulted.

An examination construction team, all of whom had previous experience with the British Columbia Ministry of Education's Student Assessment Branch, simultaneously constructed two forms of a Biology 12 examination from a common table of specifications using a pool of multiple choice items from previous examinations. A sample of students was obtained in the Okanagan, Thompson, and North Thompson areas of British Columbia. Both forms were administered to each student, as required by the test equating design (Design II (Angoff, 1971)) chosen. The data sample consisted of responses from 286 students.

The data were analyzed using a classical item analysis (LERTAP, Nelson, 1974) followed by a 2x2 order-by-form fixed effects ANOVA with repeated measures on the second factor. Item analysis revealed all items on both forms performed satisfactorily, ruling out an alternate hypothesis of flawed items being the cause of the lack of equivalence found. Results showed a significant (p<.05) difference in the means of the two forms, no
significant (p>.25) order effect, and a significant (p<.25) order-by-form interaction.

Linear and equipercentile equatings were carried out. The linear and equipercentile equatings yielded very similar results. Equating errors, judged using the conditional root mean square error of equating, was 4.86 points (9.35%) for both the equatings. Equivalency was also judged employing a graphical procedure in which the deviation of the equating function from the identity function was plotted with error bands produced using the standard error of equating. The procedure showed the two forms to be nonequivalent, particularly for the lower scoring students.

The source of the nonequivalency was investigated by separating the forms into three subtests based on the pairs of items possessing or lacking item statistics at the time of test construction. The linear equating followed by the graphical analysis was repeated for the pairs of subtests. The pairs of subtests comprised of item pairs for which difficulty (p) values were present at time of construction for one or both of the items in an item pair were found to be equivalent. In contrast, the pair of subtests comprised of items for which p values were unavailable for either item in an item pair at time of construction were found to be not equivalent.

It was concluded that the examination construction procedure in its present state cannot be relied on to produce equivalent forms. An experienced examination construction team was unable to accurately match items based on the level of
difficulty for items which did not have prior item statistics. As such, a necessary requirement for construction of equivalent forms is that item statistics be present at the time of construction.

Research Supervisor ..............................................

Dr. D. J. Bateson
# TABLE OF CONTENTS

**ABSTRACT** ........................................... ii  
**LIST OF TABLES** ......................... ix  
**LIST OF FIGURES** ............... x  
**ACKNOWLEDGEMENT** .................. xi  

## Chapter 1  
**INTRODUCTION.** ........................................ 1  
  The Grade 12 Examination Program ............... 1  
  Purpose of the Examinations .................. 2  
  Nonequivalency ................................... 3  
  Changes in Examination Assembly  
    Procedures ................................... 5  
  Statement of the Problem .................. 6  
  Definition of Terms  
    Equating ................................... 6  
    Parallel Forms ................................ 7  
    Equivalent Tests ................................ 7  
    Equivalent Scores ................................ 7  
    Comparable Scores  
    Hypotheses ................................... 8  
    Outline of Thesis  

## Chapter 2  
**REVIEW OF THE LITERATURE** .................. 11  
  Equivalence versus Comparability .................. 11  
    Parallelism ................................... 12  
      Psychological Parallelism .................. 12  
      Statistical Parallelism ..................... 13  
  Equating Designs ................................ 13  
  Equating Methods ................................ 15  
    Linear Equating ................................ 16  
      Design I Equivalent Groups Design .......... 16  
      Design II Test-Retest Design ............. 17  
      Design III Anchor Test Design ............ 19  
  Equipercen tile Equating ....................... 20  
  Equating Error ................................ 20  
  Conditional Root Mean Square .................... 21
Preliminary Classical Item Analysis ................. 47
Order of Administration ................................ 48

5 ANALYSES AND RESULTS ................................ 50

Pre-Equating Analysis .................................... 50
Statistical Properties of the Forms ..................... 50
Equating ...................................................... 51
  Linear Equating Angoff's Design II ................. 51
  Linear Equating .......................................... 52
  Equipercentile Equating ................................ 54
Comparison of the Two Equatings ....................... 57
  Differential Performance at Lower Achievement Levels .... 58
Assessment of Equivalence of Forms .................... 59
  CRMS and AAD ............................................. 59
  Standard Error of Equating ............................ 59
Nonequivalence ............................................. 61
  Source of Nonequivalence ............................... 61
  Equating of Subtests .................................... 61

6 CONCLUSIONS ............................................. 65

Summary .................................................... 65
  Purpose and Problem ..................................... 65
  Procedure ................................................. 65
  Analysis and Results .................................... 66
Limitations ............................................... 68
Conclusions ............................................... 68
Implications ............................................... 69
  Implications for Practice ............................... 69
  Implications for Future Research ...................... 70

REFERENCES .............................................. 71

APPENDIX A: Ministry Examination Construction Process 76
APPENDIX B: Materials for Construction Team .................. 79
APPENDIX C: Examination Forms ............................. 86
APPENDIX D: Request and Permission Forms . . 117
APPENDIX E: Teacher Administration Instructions . 125
APPENDIX F: Equipercenitle Equating Tables . . 129
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1989 Examination Means and Standard Deviations</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>Three Random Group Equating Designs</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>1900 Biology 12 Table of Specifications</td>
<td>35</td>
</tr>
<tr>
<td>4</td>
<td>Form A and Form B Table of Specifications</td>
<td>38</td>
</tr>
<tr>
<td>5</td>
<td>Test Administration Design and Estimated Enrollment</td>
<td>44</td>
</tr>
<tr>
<td>6</td>
<td>Summary ANOVA</td>
<td>48</td>
</tr>
<tr>
<td>7</td>
<td>Cell Means</td>
<td>49</td>
</tr>
<tr>
<td>8</td>
<td>Psychometric Properties of the Two Forms</td>
<td>51</td>
</tr>
<tr>
<td>9</td>
<td>Weighted and Pooled Means and Variances</td>
<td>52</td>
</tr>
<tr>
<td>10</td>
<td>Results of the Linear Equating</td>
<td>53</td>
</tr>
<tr>
<td>11</td>
<td>Results of Equipercentile Equating</td>
<td>56</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure

1  Linear Equating . . . . . . . . . . 54
2  Equipercntile Equating . . . . . . . . 56
3  Standard Error of Equating Total Examination . 60
4  Standard Error of Equating P Values Known for Both Items . . . . . . . . . . 62
5  Standard Error of Equating P Value Known One Item Only . . . . . . . . . . 63
6  Standard Error of Equating P Values Unknown for Both Items . . . . . . . . . . 64
ACKNOWLEDGEMENT

I would like to express my appreciation and gratitude to the many individuals and organizations that made completion of this study possible. Although all of my thesis committee members have offered insightful comments, and given me valuable support beyond what is normally required, all have given me much more. Special thanks are offered to:

Dr. David J. Bateson for first stirring my interest in educational measurement;

Dr. Harold Ratzlaff for the encouragement that lead to my entrance into the MERM area;

Dr. W. Todd Rogers, not only for his advice and support during my course work both at the University of British Columbia and later the University of Alberta, but most especially during the writing of this thesis.

This study could not have been carried out without a grant from the B. C. Ministry of Education. Personal financial assistance in the form of course fee reimbursement was given by the Kamloops District Teachers' Association. I appreciate the time and effort put forth by the students and teachers who took part in this study. I am especially indebted to the examination assembly team for the use of their time and expertise.

Finally, this thesis could not have been completed without the invaluable support of my family, and my friends and colleagues at UBC, the U of A, and in the Kamloops School District.
CHAPTER 1
INTRODUCTION

In 1983, the British Columbia Ministry of Education reintroduced provincial examinations for thirteen grade 12 subjects. While there have been, and still are many issues surrounding British Columbia governmental examinations (Anderson, Muir, Bateson, Blackmore, & Rogers, 1990; Rogers 1990; Ministry of Education, 1985; Sullivan, 1988), this study is concerned with only one of these issues. Seven years after the reinstatement, there are still concerns about the ability of the test makers to create equivalent forms that are required for use at the three testing times -- January, June, and August -- throughout the school year. In an attempt to address this concern, the Ministry has changed its procedures for examination construction. The purpose of this study was to assess the success of these new procedures.

The Grade 12 Examination Program

All Grade 12 students in British Columbia are required to write school leaving examinations in each examinable course in which they are enrolled. All grade 12 students must write at least one examination in order to graduate, either English 12 or Communications 12. Presently examinations are administered in fifteen courses: Algebra (Mathematics as of September 1990), Biology, Chemistry, Communications, English, English Literature, Francais-langue, French, Geography, Geology, German, History, Latin, Physics, and Spanish. The scores from each examination are
combined with the school awarded mark so that the examination counts 40% of the final mark for each course.

The examinations are offered on three occasions during the school year: January, June, and August. The January sittings allow students in semestered schools to write examinations for their first term courses at the appropriate time. The June sittings are the times at which students in semestered schools write their second term examinations. Students in nonsemestered schools write all of their examinations in June. Some students are granted the opportunity to write one or more examinations in August.

Purpose of the Examinations

The expressed purposes of the Grade 12 examinations are

... to insure that grade 12 students meet consistent provincial standards of achievement in academic subjects. The examination program will also ensure that graduating students from all schools will be treated equitably when applying for admission to universities and other post-secondary institutes. An additional purpose is to respond to strong public concerns for improved standards in education. (Ministry of Education, 1983, p. 6)

The stated reasons for the examinations center around a desire for consistent standards, whether the consumers of this information be the Ministry, universities, teachers, parents, or students themselves. It is this consistency that is the issue of this study.
Nonequivalency

The examinations are prepared using a common table of specifications, the assumption being that the forms produced will be equivalent. Preliminary inspection of the examinations using statistical criteria indicates there may possibly be differences between forms in some subject areas (Ministry of Education, 1989). Table 1 displays means and standard deviations of the multiple choice sections for the 1989 examinations.

Table 1
1989 Examination Means and Standard Deviations

<table>
<thead>
<tr>
<th>Subject</th>
<th>January</th>
<th>June</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>M</td>
</tr>
<tr>
<td>Alg</td>
<td>2857</td>
<td>34.0</td>
</tr>
<tr>
<td>Bio</td>
<td>2007</td>
<td>32.3</td>
</tr>
<tr>
<td>Chem</td>
<td>1234</td>
<td>32.2</td>
</tr>
<tr>
<td>Comm</td>
<td>1094</td>
<td>35.5</td>
</tr>
<tr>
<td>Eng</td>
<td>6420</td>
<td>18.3</td>
</tr>
<tr>
<td>Lit</td>
<td>577</td>
<td>19.0</td>
</tr>
<tr>
<td>Fren</td>
<td>1136</td>
<td>45.4</td>
</tr>
<tr>
<td>Geog</td>
<td>1599</td>
<td>35.9</td>
</tr>
<tr>
<td>Hist</td>
<td>1562</td>
<td>31.9</td>
</tr>
<tr>
<td>Phys</td>
<td>432</td>
<td>39.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t-test</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-2.33*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-26.88*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.28*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-9.63*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.00*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.01*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p<.05
Examination of Algebra 12 (Alg), Biology 12 (Bio), Chemistry 12 (Chem), English 12 (Eng), and Geography 12 (Geog) show little difference, less than one point, between the means of the respective January and June examinations. While the Algebra means are significantly different ($t = -2.33$, $p < .05$), the difference of 0.5 points might be considered a trivial difference. Communications 12 (Comm), English Literature 12 (Lit), French 12 (Fren), History 12 (Hist), and Physics 12 (Phys) all show significant ($p < .05$) and nontrivial differences. The equal means of the January and June forms of the examinations for the first group of subjects, Algebra, Biology, Chemistry, English, and Geography suggest that the January and June forms for a given subject are of equal difficulty and that the January and June examinee populations are of equal ability within the various subject areas. If then, the examinee populations are of equal ability, the difference in means of the second group of subjects must be due to January and June forms that vary in difficulty.

The Ministry of Education is aware of the possible lack of equivalency between forms as reflected by the following statement which appears in its Biology 12 Report to Schools (1988):

Without field test information it is not possible to ensure that the January, June, and August examinations will be of equal difficulty. (Biology 12 Report to Schools, Appendix C, p. 2)
The Ministry has attempted to address this problem through the use of a standard setting process (Ministry of Education, 1989, p. 75). Students do not receive raw score percentages as their examination marks. Instead, the province-wide distribution of examination results for a given subject is compared with the province-wide distribution of school-awarded marks for the given subject. If a difference between the two distributions occurs, the examination distribution is altered to more closely fit the school awarded marks distribution. This is done in spite of a Ministry caveat that the two marks do not measure the same skills or knowledge.¹

Changes in Examination Assembly Procedures

In an attempt to address the apparent lack of equivalency of examinations in a given school year, the Ministry implemented a new test construction procedure (see Appendix A) in 1990 (Ministry of Education, 1990). Prior to 1990, the forms for a given school year were produced separately at different times of the year, although by the same team. Beginning with the 1990 examinations, the three forms were assembled simultaneously.

¹ The August results are not discussed for two reasons. First, the Ministry of Education does not routinely release information pertaining to the August examinations as it does for the January and June sittings. Second, students writing one or more examinations in August do so either because of a previous examination failure that school year or due to some other reason acceptable to the Ministry of Education. The examinees are clearly neither a random nor a representative sample of the population of students writing examinations in any given school year. Knowledge of the August results would yield no information by which the equivalency of the three forms could be assessed.
using items that had already been developed, and, in some cases field tested.

Statement of the Problem

The purpose of this study was to empirically test the assumption that the examination forms produced by the new procedure are equivalent. The specific question addressed was:

For a given subject, will the two forms of the government examination constructed at the same time by the same team from a common table of specifications be equivalent?

It is in the interests of all the stakeholders that the exams be made as fair and consistent as possible. Not only must the results of the examinations be as consistent as possible, but it must also be demonstrated that this consistency exists. There is an "...obligation to review whether a practice has appropriate consequences for individuals and institutions, and especially to guard against adverse consequences" (Messick, 1989, p. 20). The intent of this study was to determine whether or not the procedure of simultaneous construction of the examination forms would produce the equivalent forms of an examination that would ensure consistent treatment of all examinees.

Definition of Terms

Equating

For the purposes of this study, the definition of equating proposed by Angoff (1971) was adopted:
equating, or the derivation of equivalent scores, concerns itself with the problem of unique conversions which may be derived only across test forms that are parallel, that is forms that measure, within acceptable limits, the same psychological function. (p. 563)

Parallel Forms

Equivalence is dependent upon parallelism, in fact is synonymous with it (Holmes, 1981, p. 7). Parallel tests can be defined using both statistical and psychological criteria. Gulliksen (1950) defined parallelism as follows:

In addition to equal means, variances, and reliabilities, parallel tests should have approximately equal validities for any criterion...The tests should contain items dealing with the same subject matter, items of the same format, etc. In other words, the tests should be parallel as far as psychological judgement is concerned. (pp. 173-174)

Equivalent Tests

Dorans and Lawrence (1990) describe tests as being equivalent if the equating function, the function that maps scores on one test to scores on another, is the identity function, \( y=x \), (p.253).

Equivalent Scores

Test forms which meet the criteria for parallelism are referred to as parallel forms or equivalent forms. The scores obtained from these forms are interchangeable; they can be interpreted in the same way.
Comparable Scores

If the tests are not parallel, then comparable scores may be calculated, but these scores do not function in the same fashion. For example, scores on a spelling quiz could be mapped to comparable scores on a multiplication quiz, yet no one would claim these scores measure the same construct. Such scores, while comparable, are not equivalent. Rogers and Holmes (1987) state:

{computing comparable scores} can be done irrespective of test content and carries no implication of the interchangeability of tests. (p. 4)

If two scores are merely comparable, they correspond to equal standard scores or the same rank position if the score distributions are similar. However, the scores do not carry any connotation as to the equality of the amount of the trait(s) being measured (Rogers & Holmes, 1987, p. 5). The score conversions are nonunique and are not generalizable beyond the group for which they were established (Millman & Linoff, 1964).

Hypotheses

Given the Gullicksen (1950) definition of parallel forms, the following three hypotheses for the study were formulated:

1. \( H_0 \quad \mu_a = \mu_b \)
   \( H_1 \quad \mu_a \neq \mu_b \),

where \( \mu_a \) and \( \mu_b \) are the means on Form A and Form B, respectively.
2. \( H_0 \sigma^2a/\sigma^2b = 1 \)  
   \( H_1 \sigma^2a/\sigma^2b \neq 1 \),  
where \( \sigma^2a \) and \( \sigma^2b \) are the variances of Form A and Form B, respectively.

3. \( H_0 \rho_{AA} = \rho_{BB} \)  
   \( H_1 \rho_{AA} \neq \rho_{BB} \),  
where \( \rho_{AA} \) and \( \rho_{BB} \) are the internal reliabilities of Form A and Form B, respectively.

A fourth hypothesis, based on work by Loret, Seder, Bianchini, & Vale (1974, p. 8) and Holmes (1981, p. 80), was:

4. \( H_0 \, \text{c}\rho_{AB} = 1 \)  
   \( H_1 \, \text{c}\rho_{AB} < 1 \),  
where \( \text{c}\rho_{AB} \) is the intertest correlation coefficient between Form A and Form B corrected for attenuation.

A fifth hypothesis, based on Dorans' and Lawrence's work, was:

5. \( H_0 \, F(x) = I(x) \)  
   \( H_1 \, F(x) \neq I(x) \),  
where \( F(x) \) is the equating function and \( I(x) \) is the identity function \( y=x \).
Outline of the Thesis

Chapter Two begins with an elaboration of equivalent and comparable scores discussed in relation to psychological and statistical parallelism. Data collection designs and equating methods along with associated equating error estimates are discussed. The relative merits and weaknesses of each model as applied to this study are used to select the most feasible data collection design and equating process. In Chapter Three the construction of the examination forms used in this study is described. Chapter Four includes a description of both the methodology and the preliminary data analysis, while the main analyses and the corresponding results are presented in Chapter Five. A summary of the study together with conclusions, implications for practice, and implications for future research are presented and discussed in Chapter 6.
CHAPTER 2

REVIEW OF THE LITERATURE

As described in Chapter 1, questions have been raised regarding the equivalency of the three forms of a Grade 12 examination that are administered during a given school year. Described in this chapter are various procedures that can be employed to test the equivalency of test forms. First, equivalent and comparable scores are discussed in relation to psychological and statistical parallelism and the differences between them noted. This is followed by an outline of standard data collection designs and corresponding analytical procedures for assessing the equivalence of different test forms. Various error estimates of equating are then described. The chapter concludes with the rationale for the design selected and used in the present study.

Equivalence versus Comparability

The distinction between equating and comparing is somewhat blurred due to inconsistent use in the literature (Rogers & Holmes, p. 5). As described in Chapter One, comparable scores can be produced irrespective of test content but these scores are not interchangeable and do not carry any connotation about the amount of trait measured. Equivalent scores are interchangeable and represent equal amounts of a trait.

Likewise, terms describing the mathematical process for obtaining these equivalent and comparable scores is also blurred. Angoff (1971) uses the term calibrating as the more inclusive term for the process that maps a score on one test to a score on
another test. Thus the calibration process produces comparable scores. In the more restrictive case of parallel tests, the process of calibrating is properly described as equating. The result of an equating is equivalent scores on forms that are said to be equivalent forms. The production of equivalent scores is dependent on the requirement of parallel forms being met.

**Parallelism**

Gulliksen (1950) addressed the interchangeability of forms when he suggested that "two tests are parallel when it makes no difference which test you use." (p. 11). Later Lord and Novick (1968) clarified the definition of parallel when they stated:

> ...parallel measurements measure exactly the same thing in the same scale and, in a sense, measure it equally well for all persons. (p. 48).

Parallelism can be described in both psychological and statistical terms. Both psychological and statistical parallelism must be satisfied if the forms are to be considered parallel.

**Psychological parallelism.** Psychological parallelism requires that the test forms to be equated measure the same psychological trait or function (Rogers & Holmes, p.6). Earlier, Wesman (1958) suggested, "the degree to which the conditions of parallelism are met are most closely achieved when forms are designed to be parallel" (p. 8). Since the two forms of the Biology 12 examination, the examination selected for study in this thesis, were produced simultaneously from the same table of
specifications, they met the criteria suggested by Wesman. Therefore they were judged to be psychologically equivalent.

**Statistical parallelism.** In addition to Gullicksen's (1950) statistical criteria of equal means, variances, and reliabilities, Loret et al (1974, p. 8), and Holmes (1981, p. 80) employed the intertest correlation coefficient corrected for attenuation, $cP_{ab}$, as a test for parallel forms. The basic formula for this coefficient is:

$$cP_{ab} = \rho_{ab} / \sqrt{(\rho_{aa}\rho_{bb})},$$

(1)

where

$cP_{ab}$ is the intertest correlation corrected for attenuation,

$\rho_{ab}$ is the intertest correlation, and

$\rho_{aa}$ and $\rho_{bb}$ are the internal consistency reliabilities of Form A and Form B. (Cronbach, 1971, p. 489).

A value of .95 is a commonly accepted cutoff point for equating studies (Linn, 1975, p. 206).

**Equating Designs**

Angoff (1982) presented three basic designs for use in developing equivalent or, in the case of lack of parallelism, comparable scores. These designs are shown in Table 2.
Table 2
Three Random Group Equating Designs

<table>
<thead>
<tr>
<th>Design</th>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>X</td>
<td>Y</td>
</tr>
<tr>
<td>II</td>
<td>X:Y</td>
<td>Y:X</td>
</tr>
<tr>
<td>III</td>
<td>X, U</td>
<td>Y, U</td>
</tr>
</tbody>
</table>

In Design I, the Equivalent Groups Design, two groups formed by random selection are administered two forms of the same instrument (X and Y). The Equivalent Groups Design employs random groups and assumes equality of the two groups.

Design II, the Test-Retest Design, involves the administration of both forms to one group in one order and a randomly different group in reverse order, thereby allowing for an assessment of a possible order effect. The order effect may consist of practice effects or fatigue effects. The randomization is often achieved by randomly splitting the group in two by spiralling the test booklets so that half the sample write the forms in one order while the other half write in the reverse order.

Design III, the Anchor Test Design, consists of different forms administered to random groups as in Design I but with a set of common or anchor items (labelled U in Table 1). The anchor items may be placed within the forms of the test or administered
in a separate form. Anchor items are placed on all forms of the instrument so that comparisons between performance of the two groups can be made. Use of the Anchor Test Design, Design III, reduces the need for the assumption of equivalency of the two groups required by Design I.

Equating Methods

An equating method is an empirical procedure for determining a transformation to map the scores on one form of a test onto the scores of a second parallel form of the test (in the case of parallel tests) or to map the scores of one test onto another (in the case of comparable scores). There are three methods for determining the transformation: linear equating, equipercentile equating, and item characteristic curve (ICC) equating (Petersen, Marco, & Stewart, 1982, p. 73). Linear and equipercentile equating may be carried out using observed scores (Braun & Holland, 1982, pp. 9-49) within the classical test model; ICC equating is carried out using item response models.²

Observed score equating can be carried out if the forms are equally reliable (Angoff, 1982, pp. 58-59). Examination of the internal consistency reliability estimates for the 1989 grade 12 examinations revealed that eight of the ten examinations' reliabilities differ by .02 or less (Ministry of Education, 1989, pp.

²Classical test-model theory has been usefully applied to examinee achievement and aptitude test performance for many years. Item Characteristic Curve (ICC) equating, which uses item response theory, is an alternative to classical observed score equating (Lord, 1980). The reader is referred to Hambleton (1989) for a more complete introduction to item response theory.
66-67). As the forms in this study are designed following Ministry procedures, it is likely that equally reliable forms will result. The following discussion of equating deals with the case of equally reliable forms. The reader is referred to Angoff (1971) for the case of unequally reliable forms.

Linear Equating

In linear equating it is assumed that two scores are equal if they correspond to equal standard score deviates (Angoff 1971, p. 564):

\[
\frac{Y-M_y}{s_Y} = \frac{X-M_x}{s_X}, \tag{2}
\]

where \(M_y\) and \(M_x\) and \(s_y\) and \(s_x\) are, respectively, the means and standard deviations of Form Y and Form X. The linear nature is more apparent if the equation is rewritten to give:

\[
Y = AX + B \tag{3}
\]

with coefficients A and B defined according to the equating design.

**Design I Equivalent Groups Design**

The equivalent groups design consists of one administration of Form X to one group and one administration of Form Y to another group assumed to be equivalent to the first group. The values of the coefficients A and B can be calculated according to the following equations:
A = \frac{s_y}{s_x} \quad (4)

and

B = M_y - AM_x \quad (5)


**Design II Test-Retest Design**

In the case of the test-retest design each examination form is administered to both group 1 and group 2. Therefore there are two means and two standard deviations for each group. The values for A and B can be calculated according to the following equations provided the sample sizes are equal:

\[
A = \sqrt{\frac{(s^2_{y1} + s^2_{y2})}{(s^2_{x1} + s^2_{x2})}} \quad (6)
\]

and

\[
B = 0.5(M_{y1} + M_{y2}) - 0.5A(M_{x1} + M_{x2}) \quad (7)
\]


However as pointed out in Glass & Hopkins (1984), if the sample sizes are not equal, then this must be taken into account when computing values for A and B. The equations then used are:

\[
A = \sqrt{s^2_y/s^2_x} \quad (8)
\]

and
\[ B = \frac{\left( n_{y2}M_{y2} + n_{y1}M_{y1} \right)}{\left( n_{y2} + n_{y1} \right)} - \frac{\left( A \right) \left( n_{x2}M_{x2} + n_{x1}M_{x1} \right)}{\left( n_{x2} + n_{x1} \right)}, \]  

(9)

where

\( s^2_y \) is the weighted average variance on Form Y,

\( s^2_x \) is the weighted average variance on Form X,

\( M_{y1} \) is the mean on Form Y of the group of examinees who wrote Form X first,

\( M_{y2} \) is the mean on Form Y of the group of examinees who wrote Form X second,

\( M_{x1} \) is the mean on Form X of the group of examinees who wrote Form X first,

\( M_{x2} \) is the mean on Form X of the group of examinees who wrote Form X second, and

\( n_{x1}, n_{x2}, n_{y1}, \) and \( n_{y2} \) are the respective sample sizes (Angoff, 1971, pp. 574-575).

The weighted average variances are calculated with the equations:

\[ s^2_y = \frac{\left( n_{y1} - 1 \right)s^2 y_1 + \left( n_{y2} - 1 \right)s^2 y_2 + n_{y1}(M_{y1} - M_Y)^2 + n_{y2}(M_{y2} - M_Y)^2}{n_y - 1} \]  

(10)

and

\[ s^2_x = \frac{\left( n_{x1} - 1 \right)s^2 x_1 + \left( n_{x2} - 1 \right)s^2 x_2 + n_{x1}(M_{x1} - M_X)^2 + n_{x2}(M_{x2} - M_X)^2}{n_x - 1}, \]  

(11)
where $s^2_{x1}$, $s^2_{x2}$, $s^2_{y1}$, and $s^2_{y2}$, are the variances of the respective subsamples, and with other variables defined as before (Glass & Hopkins, 1984, p. 53).

Angoff (1971) describes a variation of these procedures (p. 575). The data on Form X for the two half groups are combined and likewise the data on Form Y for the two half groups. The equating is then carried out using equation 4 and equation 5.

**Design III Anchor Test Design**

The anchor test design consists of a set of common items that can be described as forming a test U. Tests X and Y are then regressed on U to form estimates for the mean ($\mu_x$ and $\mu_y$) and standard deviation ($\sigma_x$ and $\sigma_y$). The values for A and B can be calculated according to the following equations:

$$A = \frac{\sigma_y}{\sigma_x}$$

and

$$B = \mu_y - A\mu_x,$$  \hspace{1cm} (13)

where the estimated means are:

$$\mu_x = M_{x1} + b_{xu1}(\mu_u - M_{u1})$$  \hspace{1cm} (14)

$$\mu_y = M_{y2} + b_{yu2}(\mu_u - M_{u2}),$$  \hspace{1cm} (15)
and the estimated standard deviations are:

\[ \sigma_x = \sqrt{\sigma^2_{x1} + b^2_{xu1}(\sigma^2_u - \sigma^2_{u1})} \]  

(16)

\[ \sigma_y = \sqrt{\sigma^2_{x2} + b^2_{yu2}(\sigma^2_u - \sigma^2_{u2})} \]  

(17)

where \( b_{xu1} \) and \( b_{yu2} \) are the regression coefficient X on U and Y on U for the respective subsamples. (Angoff, 1971, pp. 564-577).

**Equipercentile Equating**

Equipercentile equating is based on a definition of equivalent scores (Flanagan, 1951; Lord, 1950) that considers two scores equivalent if their corresponding percentile ranks are equal. As such, equipercentile equating makes no assumptions about the score distributions for the two forms. The procedure first requires the computation of percentile ranks for the raw score distributions. These percentile ranks are plotted against the raw score. Smoothing of the line may be done by hand although computer programs that simulate the smoothing process have been developed (Lindsay and Prichard, 1971). Scores from the smoothed curves are plotted against one another and this new curve is then smoothed. Score conversions are taken from this final smoothed curve.

**Equating Error**

Three error measures are considered when assessing the degree of equivalence as determined by an equating method. The first two, the conditional root-mean-square (CRMS) and the
average absolute difference (AAD) are used to judge the equivalence between scores with the sample(s) used. The third, the standard error of equating ($s_y^*$), is used to assess the degree of sampling error in the estimates across repeated replications.

**Conditional Root Mean Square**

The CRMS is a measure of the fit between an observed score on X and the equivalent score on X derived from a knowledge of a score on Y. The formula used to calculate CRMS is:

$$\text{CRMS} = \sqrt{\frac{\sum(X-Y^*)^2}{n-1}}, \quad (18)$$

where $X$ is the score on Form X, $Y^*$ is the score derived from Form Y, and $n$ is the number of subjects writing both Form X and Form Y (Bianchini & Loret, 1974, p. 158).

This error is an estimate of "the degree to which a score read from the equating table would differ from the score a pupil would have earned had he been given the equated test" (Jaeger, 1973, p. 7). Holmes (1981) judged discrepancies of 12 points or more ($M=100$, S.D.=15) as an indication of nonequivalency (p. 137). In a practical sense, the magnitude of the error as it pertains to test interpretation can be considered an indicator of equivalency (Holmes, p.136). As well, the CRMS can serve as an indicator of the suitability of the method of equating.
The magnitude of the conditional root mean square can be compared to another similar error estimate in an attempt to assess equivalency of the forms and the interpretability of the individual scores. The standard error of measurement (SEM) "can be viewed as the standard deviation of the discrepancies between a typical examinee's true score and the observed scores over an infinite number of repeated testings" (Crocker & Algina, 1986, p. 150).

The Average Absolute Difference

The AAD is defined as:

$$\text{AAD} = \frac{\sum |X - Y^*|}{n}$$  \hspace{1cm} (19)

where $X$ is the score on Form $X$,
$Y^*$ is the score derived from Form $Y$ and,
n is the number of pairs of scores.

Like the CRMS, the AAD provides an indication of the degree to which individuals achieve the same score on two forms to be equated.

Standard Error of Equating

Unlike the CRMS and the AAD, which are indicators of the similarity of an individual's scores on the two forms, the standard error of equating, $s_{y^*}$, is a measure of error in the equating procedure itself. The standard error of equating ($s_{y^*}$) "reflects the degree to which the equating results would vary if the same
equating procedure and method were applied to different representative samples of pupils" (Jaeger, 1973, p. 7). Unlike the CRMS and the AAD, which are defined as above regardless of equating design used, the formula for $s_y^*$ depends upon the design used and the equating method employed.

**Linear Equating**

**Design I.** For Design I, the Equivalent Groups Design, the error is estimated by the equation:

$$s_y^* = \sqrt{2s_y^2(z_x^2 + 2)/n},$$  \hspace{1cm} (20)

where $s_y^*$ is the standard deviation of the error estimate,

$s^2_y$ is the variance of the scores on Form Y,

$z_x$ is a standard score on Form X, and

$n$ is the total number of examinees in both groups (Angoff, 1971, p.570\(^3\)).

**Design II.** For Design II, the Test-Retest Design, the appropriate formula is:

$$s_y^* = \sqrt{s^2_y(1-r_{xy})[(z_x^2(1+r_{xy})+2)/n]},$$  \hspace{1cm} (21)

where $s^2_y$ is the variance of the Form Y scores,

$r_{xy}$ is the correlation between the Form X and Form Y scores.

\(^3\)Braun and Holland (1982, pp.32-36) discuss the more general case of unequal group size and nonnormal score distributions.
scores, 
\[ z_x \] is the standard score for each Form X score, and 
\[ n \] is the number of examinees (Angoff, 1971, p.575).

**Design III.** For Design III, the Anchor Test Design, the appropriate formula is:

\[
s_{y*} = \sqrt{\frac{2s_y^2(1-r^2)[(z_x^2(1+r^2)+2)/n]}, \quad (22)
\]

where \( s_y^2 \) is the estimated variance of the Form Y scores, 
\( r \) is the estimated correlation between the estimated Form X and estimated Form Y scores, 
\( z_x \) is the standard score for each Form X score, and 
\( n \) is the total number of examinees (Angoff, 1971, p.577).

**Equipercentile Equating**

Lord (1982), demonstrated that the magnitude of standard errors of linear equating was one half that of equipercentile error in the middle of the score range and comparatively smaller at the extremes (p. 173). Even though Lord derives standard errors of equipercentile equating for Design I and Design II, he concludes that linear equating is clearly the method of choice (p.174). Given Lord's strong preference and, as will be reported in Chapter 5, only the equipercentile standard error of equating for Design I is discussed here.

**Design I.** For Design I, the Equivalent Groups Design, the error is estimated by the equation:
\[ s_{y'} = \sqrt{s_{y}^2pq(n_x^{-1} + n_y^{-1})/\phi^2} \]  \hspace{1cm} (22)

where \( s_{y}^2 \) is the estimated variance of the Form Y scores,
\( p \) is the proportion of scores below a score, \( x_0 \),
\( q \) is \( 1-p \),
\( \phi \) is the standard normal density at the unit normal below which \( p \) of the cases fall,
\( n_x \) is the number of examinees writing Form X, and
\( n_y \) is the number of examinees writing Form Y, (Petersen, Kolen & Hoover, 1989, p. 251)

**Design II.** For Design II, the Test-Retest Design, the appropriate formula and related discussion is found in Lord (1982).

**Design III.** For Design III, the Anchor Test Design, the appropriate formula and related discussion is found in Jarjoura & Kolen (1985).

**Use of Standard Error of Equating**

The standard error of equating \((s_{y'})\) was used to judge the equivalency of the two forms in the Anchor Test Study. Linn (1975) reported errors of less than one score point were found in the Anchor Test Study (p. 207).

Dorans and Lawrence (1990) used this error \((S_{y'})\) to place a heuristic confidence band around the equating function (p. 247). They suggested that the null hypothesis for equating be that the equating function equals the identity function (p. 253). They then
suggested placing bands of plus or minus two standard errors of equating ($\pm 2S_y^*$) around the equating function to define the region in which the identity function must fall if the tests are judged to be equivalent at the .05 level of significance. To present this information, Dorans and Lawrence suggested using a graph with the ratio of the difference between predicted score and observed score divided by the standard error of equating ($s_y^*$) plotted on the ordinate and the observed score of a test plotted on the abscissa. Forms were judged equivalent if the plotted function does not exceed bands of $\pm 2$ units.

While attractive due to its ease of interpretation, the graph of this ratio versus raw score has two flaws. First, the standard error of equating is not constant throughout the range of scores, but this is not apparent using this graph since the standard error of equating is not plotted. Second, the plot of the equating function is distorted, giving the appearance of being least equivalent near the mean of the scores. Both these shortcomings are corrected by first using bands formed by $\pm 1.96S_y^*$ which are placed around the line $y=0$. The line $y=0$, is equivalent to the difference between the equating function and the identity function if the forms are perfectly equivalent. Forms are judged as equivalent if the plot of difference between the equating function and the identity function falls within the standard error bands. This plot accurately displays both the nature and magnitude of the equating function and the standard error of equating.
Magnitude of the Standard Error of Equating

Sample size, equating design and equating method all affect the magnitude of the standard error of equating. Within a given equating method, sample size and design interact to determine the magnitude of equating error. With Design I, large sample sizes may be needed to reduce the standard error of equating. For example, Dorans and Lawrence (1990) used two random groups of 48,639 and 43,845 subjects in an equating study in which the statistical equivalence of nearly identical test editions of the Scholastic Aptitude Test (SAT) was investigated. Angoff (1971) found that for a test-retest reliability of .80, ten times the sample size is needed to reduce the standard error of equating Design I to the standard error of equating Design II (p. 575). With Design III, the anchor test scores are used to reduce error caused by group differences, whether the differences result from random or non-random group assignment. Continuing his earlier example, Angoff states that when the correlation between the anchor test and the forms to be equated equals .87, the equating error is one-fourth as large at the mean as it would be using Design I. The sample size using Design I would need to be four times that of the Design III sample if the same magnitude of error is desired.

Given a set sample size and set equating design, the method of equating also influences equating error. As described earlier, the error of equipercentile equating is generally larger than that of linear equating throughout most of the score range.
Selection of Data Collection Design

Since this study will use forms as similar as possible to a British Columbia Grade 12 Biology Examination, the content area to be tested includes the entire core curriculum which is taught throughout the course. Also, since students should be given the opportunity to learn the material to be tested, the testing should take place as near to the end of a school year. However, teachers, wary of the pending examinations, express concern about time that can be spared for a research project. It was therefore decided, that given limited time for testing and in light of cost considerations, Design II, the Test-Retest Design, (see Table 2) would be employed to collect the data. Further, these restrictions were the reasons that two rather than three test forms would be constructed at one time for a single subject, Biology 12.

Selection of Equating Model

Linear equating is appropriate if the X and Y distributions differ only in the means and standard deviations (Braun & Holland, 1982, p. 17). As it is totally analytical and free from smoothing errors, it is the preferred method if the differences between the shapes of the score distributions can be regarded as trivial. Dorans and Lawrence (1990) support the choice of the linear method arguing that because the null hypothesis for parallel forms is a linear identity function, it does not make sense to use equipercentile equating if forms are expected to be nearly parallel (p. 253).
If a linear equating function cannot be assumed, then equipercentile equating can be used. However, the lack of assumptions about the score distributions and subsequent wider applicability has a price. The error of equipercentile equating is larger than that of linear equating throughout the score range (Lord, 1982, pp. 173-174). This makes it less powerful in detecting nonequivalency between forms (Dorans & Lawrence, 1990, p. 253). Ceiling and floor effects allow equipercentile equating to be used when a linear equating produces equated score outside of the test score range. This occurred in the Anchor Test Study (Bianchini & Loret, 1974).

Skaggs & Lissitz (1986) suggest that no single method of equating is consistently superior to others. If tests are reliable, nearly equal in difficulty, and samples are nearly equal in ability, then most linear, equipercentile, and IRT models will produce satisfactory results (p. 523). For small sample sizes (n<1000), it is unclear which of these methods yields the best results. Earlier, Kolen and Whitney (1982), using the Test of General Educational Development in an equating study with a sample size of 200, found that linear and Rasch (a one-parameter logistic item response model) equating methods produced the acceptable results for horizontal equating of achievement tests, while equipercentile and three-parameter logistic (item response model) equating methods produced unacceptable results.

---

4Equating tests of similar difficulty is referred to as horizontal equating.
The examination forms for this study were simultaneously assembled according to the specifications of a common specification table. The intent of the new Ministry procedure is to construct equivalent forms. There should be no difference in difficulty. Equally reliable test forms would also be expected since most examination pairs are already equally reliable. A test-retest design is being used; there should be no difference in ability of the examinees from one testing day to the next. Under these conditions, linear equating, (Petersen, Cook, & Stocking, 1983) using linear method for equally reliable tests (equations 3-6), appeared to be the most appropriate method to follow. However, given the uncertainty in the literature about which method is superior, it was also decided that both Angoff’s linear procedure for Design II in which the data are pooled across order of administration, and equipercentile equating using pooled data would be performed. If the linear equating using pooled data proves feasible, only this method will be retained.
CHAPTER 3

EXAMINATION CONSTRUCTION

As described in Chapter 1, beginning with the 1989/90 school year the three forms of each course examination are to be constructed at the same time in an attempt to produce a greater degree of equivalency among the three forms. The degree to which this desired outcome is met was tested in this study. Due to restrictions of time for testing, two rather than three test forms were constructed at one time for a single subject, Biology 12. The construction of these two forms is described in the present chapter.

Content, Materials, and Assembly Team

Choice of Subject

The choice of Biology 12 was based on three factors. First, there was a large enrollment in this course. The large enrollment means a large potential sample size representing a wide cross section of Grade 12 students. Second, there was a sufficiently large number of multiple choice items available that would be suitable for use with the 1990 Table of Specifications adopted for this study. Finally, an intact experienced team of school teachers was available to assemble the Biology 12 forms from the pool of available items.

Number of Forms

Two forms, each containing 52 items, were produced. Producing three forms, as is the intention of the full Ministry examination program, would have required a larger sample of both
items and people as well as increased testing time, all factors which could not be accommodated in the present study. The forms were administered to students near the end of the school year to better approximate the actual time of writing. Teachers, while willing to cooperate, were willing to administer only two forms. However, if the two forms for this study, which were constructed following procedures set forth by the Ministry of Education, were found to be equivalent, then it would be reasonable to expect the process would be capable of yielding three equivalent forms. Contrarily, if the two forms were found not to be equivalent, then it would be impossible that three forms would be.

**Type of Item**

In 1987, 1988, and 1989, the provincial Biology 12 examination has consisted of approximately 50 multiple choice and 5 supply items. These items formed the pool of field tested items that were used for construction of the two forms administered in the present study. Since copies of previous examinations may be obtained from the Ministry, it is likely that each student would be familiar with at least some of the items in this pool, particularly since many teachers often use these items as review material. This creates a potential problem. An item that is recognized by an examinee because of previous experience with that item would be more likely to be answered correctly than if the item were new to the examinee. If two examination forms are created to be of equal difficulty using previous item difficulty statistics, yet one form contains more items that are recognized by the examinees sitting
the two forms, this form will appear easier due to the recognized items. Since the items to be used were to be selected from a pool of items available to the examinees, the greater number of multiple choice items than supply items provided a larger pool from which to select the items for the two forms to be constructed. This greater number of multiple choice items means that the likelihood of students recognizing a multiple choice item would be lower in comparison to the likelihood of recognizing a supply item. More importantly, the frequency of item recognition over the entire examinee population should be randomly equal on the two forms. With the much smaller sample of supply items, recognition of items is more likely to occur and an unequal number of items recognized on each form is more likely to result. Thus, given the limited time for testing and the greater number of potential multiple choice items, the decision was taken to include only multiple choice items on the two forms.

**Content Tested**

The Biology 12 curriculum consists of a core area and six options. The provincial examination for Biology 12 reflects this structure; students must write a set of core items and then select two of the six options. Assuming for the sake of argument that each option is chosen by an equal number of examinees, then each option is attempted by approximately 1/3 of the total examinees. If it were essential to obtain item statistics on the optional items equivalent in precision to the statistics for items included in the core area, a sample of examinees three times as large as
that for the core items would be needed. Consequently, given the restrictions mentioned earlier, the sample of items for the tests used in this study was restricted to the core content only.

**Item Bank**

The item bank available consisted of 288 multiple choice items in the core topics of 'cells' and 'humans' that were used on the 1987, 1988, and 1989 Biology 12 examinations. However, only the 184 items from the 1987 and 1988 examinations had any available item statistics, and only the percentage of students who chose each option was available. Statistics for the 1989 items were not available at the time of examination assembly.

**Examination Assembly Team**

The team of five teachers responsible for the construction of the two forms of the Biology 12 examination used in this study were members of a team currently involved in the Ministry's item banking project for Biology 12. The five teachers each had received training and gained experience in item writing, item classification, and mini-test construction. The five teachers had established a stable working relationship through working on the item banking project for over a year. The team chairperson was a teacher who, for several years prior to the item banking project, had served on the Ministry's Biology 12 examination production team. As such, they were comparable in knowledge and experience to the Ministry of Education teams that construct the Biology 12 examinations. Due to illness, one member of the team later dropped out of the project.
Table of Specifications

The specifications table used, presented in Table 3, was adapted from the 1990 Table of Specifications for Biology 12.

Table 3
1990 Biology 12 Table of Specifications

<table>
<thead>
<tr>
<th>Topic/subtopic</th>
<th>Cognitive level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>K*</td>
</tr>
<tr>
<td>Methods &amp; principles</td>
<td>5 ITEMS</td>
</tr>
<tr>
<td>I</td>
<td></td>
</tr>
<tr>
<td>Experimental Design</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td></td>
</tr>
<tr>
<td>Homeostasis</td>
<td></td>
</tr>
<tr>
<td>Cells</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td></td>
</tr>
<tr>
<td>Cell compounds</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td></td>
</tr>
<tr>
<td>Ultrastructure</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td></td>
</tr>
<tr>
<td>Ultraprocesses</td>
<td></td>
</tr>
<tr>
<td>VI</td>
<td></td>
</tr>
<tr>
<td>Photosynthesis</td>
<td></td>
</tr>
<tr>
<td>Subtotal</td>
<td>5</td>
</tr>
<tr>
<td>Humans</td>
<td>10</td>
</tr>
<tr>
<td>VII</td>
<td>Cells, Tissues and</td>
</tr>
<tr>
<td>VIII</td>
<td>Digestive</td>
</tr>
<tr>
<td>IX</td>
<td>Circulatory</td>
</tr>
<tr>
<td>X</td>
<td>Nervous</td>
</tr>
<tr>
<td>XI</td>
<td>Excretion &amp;</td>
</tr>
<tr>
<td></td>
<td>Respiration</td>
</tr>
<tr>
<td>XII</td>
<td>Endocrine</td>
</tr>
<tr>
<td>Subtotal</td>
<td>15</td>
</tr>
<tr>
<td>TOTAL</td>
<td>15</td>
</tr>
</tbody>
</table>

*K = Knowledge  U/A = Understanding/Application  
HMP = Higher Mental Process
The specifications table includes the core topics of 'Cells' and 'Humans' and reflects the restriction to multiple choice items. It should be noted that each item was classified by a two dimensional grid of "topic" and "cognitive level". The British Columbia Ministry of Education uses Bloom's taxonomy (Bloom, 1956) collapsed to three levels: knowledge (K), understanding/application (U/A), and higher mental processes (HMP). This table contains the information given to the examination assembly team to form a basis for their selection of items.

The numbers within the cells are the number of items for each topic by cognitive level cell. A total of five items from the 'Cells' and 'Humans' topics are also included by the Ministry in the category 'Methods and Principles'. Since these five items are already included within the 'Cells' or 'Humans' topics, they are not included again in the item subtotals or item total.

Examination Construction Process

The Ministry document titled 'FLOWCHART and TIMELINES' for the Production of the 1991 Examinations DRAFT' (see Appendix A) describes the entire new procedure to be followed to construct the forms of the examination. Only the stages that are directly related to this study are discussed below. A package consisting of instructions, the examination specifications table, items, and item option p-values (when available) was sent to all team members. The materials with the exception of copies of the items are included in Appendix B.
Examination Construction

Prior to meeting on the examination assembly day the team leader assigned a topic area and corresponding subtopics (see Table 3) to each member of the team. Each team member was asked to confirm through personal professional judgement the curricular validity of the items referenced on the basis of previous years classification for the subtopics for which they were responsible. Their judgements were then discussed at the beginning of the meeting to achieve consensus on the curricular validity. Only items which the team unanimously judged to measure the subtopic were retained.

Allocation of Items to Forms

The published table of specifications (Table 3) gives subtotals for only the very broad topics of 'Cells' and 'Humans' by cognitive level. Following identification of suitable items, the assembly team determined the number of items to be included for each subtopic by cognitive level. The results of this assignment are shown in Table 4.

To actually assemble the test forms, the team divided into two pairs, each pair responsible for one of the topic areas. Working within each subtopic by cognitive level subcell, each pair first categorized each retained item according to difficulty level. The proportion of examinees answering each item correctly (item p values) were available for the 1987 and 1988 items. The difficulty levels for the 1989 items were established subjectively by each pair of team members.
### Table 4
Form A and Form B Table of Specifications

<table>
<thead>
<tr>
<th>Topic/subtopic</th>
<th>Cognitive level</th>
<th>Form A, B</th>
<th>Form A, B</th>
<th>Form A, B</th>
<th>Subtotal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>K</td>
<td>U/A</td>
<td>HMP</td>
<td></td>
</tr>
<tr>
<td>Form</td>
<td></td>
<td>A, Ba</td>
<td>A, B</td>
<td>A, B</td>
<td></td>
</tr>
<tr>
<td>Methods &amp; principles</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I Experimental Design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II Homeostasis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cells</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>III Cell compounds</td>
<td></td>
<td>1,1</td>
<td>2,3</td>
<td>1,1</td>
<td></td>
</tr>
<tr>
<td>IV Ultrastructure</td>
<td></td>
<td>1,1</td>
<td>3,3</td>
<td>0,0</td>
<td></td>
</tr>
<tr>
<td>V Ultraprocesses</td>
<td></td>
<td>3,3</td>
<td>3,2</td>
<td>2,2</td>
<td>19,19</td>
</tr>
<tr>
<td>VI Photosynthesis</td>
<td></td>
<td>0,0</td>
<td>3,3</td>
<td>0,0</td>
<td></td>
</tr>
<tr>
<td>Subtotal</td>
<td></td>
<td>5,5</td>
<td>11,11</td>
<td>3,3</td>
<td></td>
</tr>
<tr>
<td>Humans</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VII Cells, Tissues</td>
<td></td>
<td>1,1</td>
<td>0,0</td>
<td>0,0</td>
<td></td>
</tr>
<tr>
<td>VIII Digestive</td>
<td></td>
<td>3,3</td>
<td>4,4</td>
<td>0,0</td>
<td></td>
</tr>
<tr>
<td>IX Circulatory</td>
<td></td>
<td>2,2</td>
<td>4,4</td>
<td>1,1</td>
<td>33,33</td>
</tr>
<tr>
<td>X Nervous</td>
<td></td>
<td>3,3</td>
<td>4,4</td>
<td>0,1</td>
<td></td>
</tr>
<tr>
<td>XI Excrete &amp; Respire</td>
<td></td>
<td>2,2</td>
<td>4,4</td>
<td>1,0</td>
<td></td>
</tr>
<tr>
<td>XII Endocrine</td>
<td></td>
<td>1,1</td>
<td>3,3</td>
<td>0,0</td>
<td></td>
</tr>
<tr>
<td>Subtotal</td>
<td></td>
<td>12,12</td>
<td>19,19</td>
<td>2,2</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>17,17</td>
<td>30,30</td>
<td>5,5</td>
<td>52,52</td>
</tr>
</tbody>
</table>

*a Form A number of Items, Form B number of items.*
Once this had been done, items within the same subcell were paired according to similar difficulty. One item of each pair was then randomly assigned to each form. Three different situations occurred during the matching of items. First, both items in a pair had p value information. There were 24 such pairs. Second, an item with known p value was matched with an item whose difficulty was judged subjectively by the test assembly team. There were 17 such pairs; 8 items with known p values were placed on Form A and 9 items on form B. The remaining 11 pairs were formed with no prior p value information about either item. In four instances, a matched pair of items within a subtopic area could not be formed. In these cases, items from different subtopics within the same topic area, at the same cognitive level, and with similar difficulty levels were combined for assignment purposes. The final number of items selected for each cell is reported in Table 4. As shown in Table 4, the number of items classified as either K or U/A is greater than the intended number shown in Table 3. This difference is due to the lack of availability of matched item pairs within some subtopics at the HMP cognitive level.

Items were placed on each form according to the order given in the table of specifications (see Table 4). Draft copies of each form were prepared. These draft forms were checked by the team for duplication of items or the possibility that information in one item could be used to correctly answer another item.
Assessment of Comparability

The composition of the two forms is shown in Table 4 and a copy of each examination form is provided in Appendix C. The item subtotals for both the Topic by Cognitive Level and Subtopic by Cognitive Level are identical on both forms. The only differences between the structures of the two test forms are in the 'Cells by U/A' cell: 'cell compounds' and 'ultraprocesses' subcells and the 'Humans by HMP' cell: 'nervous' and 'excrete and respire' subcells. These differences were not expected to create any differences in examinee performance between the two forms. The mean p value for the items of known p value was 0.64 for each form.

Production of Examination Booklets

The final forms of the examination were printed by a professional printer using camera ready copy provided by the researcher. Prior to printing, the camera ready copy was reviewed by a Biology 12 teacher who was not on the examination assembly team. This was done to produce an independent check of the technical aspects of the form; the clarity and accuracy of the diagrams; correctness of spelling; spacing; and item formatting.

---

5Form A and Form B were initially referred to as Form I and Form II respectively. The designations, Form I and Form II, are used consistently throughout the appendices.
CHAPTER 4
METHODOLOGY AND PRELIMINARY ANALYSES

Chapter Four consists of two major sections. In the first section, the procedure used to select the sample is described. This is followed by a description of the test administration and scoring procedures. The second section contains a description of the preliminary analysis of the data. The purposes of the preliminary analysis were to determine the characteristics of the sample and the characteristics of the two forms that were relevant to the choice of equating methods used to test the equivalence of the two forms.

Sample Selection

Sample Size

The desired sample size was set at 500. For forms that should be very nearly parallel, one might want to detect a one mark difference between the means using an alpha level of .05. Using these values, and in light of the results of previous Biology 12 examinations (standard deviation 10 for approximately 50 items), the t-test between independent groups indicated a sample size of 400 would provide the statistical precision necessary to test the equivalence of the two forms using the test-retest design with independent groups (Angoff, 1971, pp. 58-59).

School Selection

Sample schools were selected in two stages. First, officials in four school districts in the Okanagan, Thompson, and North Thompson geographical areas in British Columbia were asked for
permission to approach Biology 12 teachers in their districts to request that they and their students participate in the study (see Appendix D). These four districts were selected so that the combined Biology 12 course enrollment was greater than 700, thereby allowing for possible nonresponse at the district, school, and student levels. Initial verbal inquiries of teachers in these districts yielded a high expressed willingness to participate; therefore an initial target sample size of 700 was judged adequate.

Permission was granted in three districts. Officials in the fourth district indicated that they did not wish to have students in their district approached. No reason for refusal was given nor was one requested. Attempts to replace this district with a district of similar population size in the same region were unsuccessful due to the lateness of the request. Consequently, the potential sample size at the end of stage 1 was approximately 600.

In the second stage, a letter was sent to all Biology 12 teachers within participating districts. This letter contained a brief explanation of the process of the study and a request that their Biology 12 classes take part (see Appendix D). Of the 16 schools in which these teachers taught, teachers in 15 schools initially agreed. The single Biology 12 teacher in the remaining school explained that because of time considerations his classes would not be able to participate. All Biology 12 teachers in each of the remaining 15 sample schools agreed to administer the two forms of the examination to all students in their Biology 12
classes. Based on teacher enrollment figures the total number of potential subjects was 597. This figure of 597 is somewhat high as teachers who did not have exact figures readily available when contacted were asked to overestimate. As this estimate was used to determine the number of examination forms to be printed, overestimation was more desirable than underestimation.

The request for permission letters along with administration instructions and the 'Request for Ethical Review' form were sent to The University of British Columbia Behavioral Sciences Screening Committee For Research and Other Studies Involving Human Subjects. A copy of the Certificate of Approval received is provided in Appendix D.

**Examination Administration**

Students were tested by their teachers during the last four weeks of the school year. The test administration design and the estimated enrollment as reported by the teachers for their respective schools are shown in Table 5. As shown, Form A was scheduled first in 5 schools and Form B first in 10 schools. While there are advantages to interleaving the forms within classes, students within a given class were all given the same form at the same time. Further, all classes within a school were given the forms in the same order and on the same day. The concern that students would share information between test sittings was the reason for both practices.
Teachers were asked to allow up to a maximum of 5 school days between administrations and not to administer both forms at one sitting.

Table 5
**Test Administration Design and Estimated Enrollment**

<table>
<thead>
<tr>
<th>School</th>
<th>Order</th>
<th>AB</th>
<th>BA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AB</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>BA</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>BA</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>BA</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>BA</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>BA</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>BA</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>AB</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>BA</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>AB</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>BA</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>BA</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>BA</td>
<td>73</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>AB</td>
<td>130</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>AB</td>
<td>45</td>
<td></td>
</tr>
</tbody>
</table>

Total Enrollment 291 306 n=597

Beyond these instructions teachers were allowed to schedule the examinations at their convenience. This was necessary given the time of year at which the testing was completed. During the last four weeks of school there is a general increase in the number of activities that impinge upon the time spent in the classroom. However, despite the existence of these activities, it was
necessary to administer the tests toward the end of the school year as the content examined required the students to have completed the core sections of the Biology 12 curriculum.

Teachers were asked not to inform the students following the first test administration that they (the students) would be tested with the second form. Further, they were asked not to discuss content relevant to the examinations during the interval between testings.

The necessary number of test forms (see Appendix D) and instructions (see Appendix E) were sent to each cooperating school during the first week of May. Students were allowed 40 minutes to write each test form. Students recorded their responses on an NCS General Purpose Answer Sheet. The same sheet was used on both occasions with Form A answers recorded on side 1 and Form B answers on side 2. Following both test administrations, teachers collected the answer sheets and returned the sheets to the researcher for scoring and analysis.

**Scoring and Data Preparation**

The completed answer sheets were processed using the optical scanner maintained by the Educational Measurement Research Group (EMRG) at the University of British Columbia. The tests were scored and an item analysis was performed using LERTAP (Nelson, 1974). Scoring and statistical analyses were completed using various SPSS-X (SPSS Inc., 1988) computer programs. The analyses were performed on the Amdahl 470/V8.
Preliminary Analyses

Response Rate

Altogether, answer sheets were received for 312 students, or 52.3% of total teacher estimated enrollment. After initial agreement to participate, teachers at two schools opted out of the study, reducing potential sample size by 175 students. The reason given for opting out was lack of time. An additional 73 answer sheets from one school were lost in the mail. The mail loss, the two schools opting out, and the 312 completed forms account for 560 of the estimated 597 students. The difference of 37 (6.19%) is likely due to a number of causes. Students who had written one or neither of the two forms should not have had an answer sheet returned. Class enrollments may decrease slightly due to the student dropping Biology 12 near the end of the school year.

From the answer sheets received, those from one school with seven students were deleted due to failure to follow instructions. The students in this school were allowed to begin their second form once they had completed their first form rather than waiting until the next class, the stipulated minimum amount of time between testings. Eight students who, due to absence, completed only one of the two forms were removed from the sample. Lastly, 11 students who failed to see Item #52 Form A were removed. This problem will be discussed in detail in the preliminary
classical item analysis. Thus, the final number of usable Form A and Form B pairs was 286.

The loss of 248 students due to schools opting out and mail loss reduces the sample size considerably. The schools were urban schools that enrolled an estimated two to six classes of Biology 12. The 1990 Biology 12 examination means and standard deviations for the two schools that opted out revealed little difference between either of these schools and the means and standard deviations for the district. Based on these results there is no suggestion of bias due to the loss of these two schools.

Preliminary Classical Item Analysis

The student responses were analyzed using LERTAP (Nelson, 1974). Inspection of the percentage of students who responded to each item in each form revealed that the last item in Form A had a larger nonresponse than the preceding 51 items. Twelve subjects (4.0%) of the sample did not respond. This nonresponse rate is substantially higher than the mean item nonresponse rate of 0.1% (S.D=0.2%) for the remaining items.

The value of the point biserial, 0.00, for the nonresponse category indicates no correlation between the ability of the subjects and the lack of response to this item. Further Item #52 appeared alone on the second to the last page while the last page was blank. The placement of the item on the form, together with the item results, suggested that the 12 students likely did not see this item.
However, of the 12 students, one subject also omitted Item #51. Thus, in contrast to the other 11 students, it seemed plausible that this particular student omitted Item #52 for reasons other than poor placement of the item. Consequently, only 11 students were deleted from the sample.

**Order of Administration**

The order of administration was tested with a 2x2 (order of administration-by-test form) fixed effects ANOVA with repeated measures on the second factor. The results are summarized in Table 6.

As revealed in Table 6, there is no significant examination form order of administration effect ($F=1.33$, $p<.25$) but there is a significant examination form effect ($F=25.18$, $p<.05$).

**Table 6**

**Summary ANOVA**

<table>
<thead>
<tr>
<th>Part of model</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Subjects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Order</td>
<td>1, 284</td>
<td>209.51</td>
<td>1.33</td>
<td>.25</td>
</tr>
<tr>
<td>Subject(order)</td>
<td></td>
<td>156.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within subjects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Form</td>
<td>1, 284</td>
<td>313.56</td>
<td>25.18</td>
<td>.00</td>
</tr>
<tr>
<td>Order*Form</td>
<td>1, 284</td>
<td>61.45</td>
<td>4.93</td>
<td>.03</td>
</tr>
<tr>
<td>Form*Subjects(order)</td>
<td>1, 284</td>
<td>156.99</td>
<td>12.61</td>
<td>.00</td>
</tr>
<tr>
<td>Error</td>
<td></td>
<td>12.45</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

There is also a significant examination form by administration order interaction ($F=12.16$, $p<.25$). Of interest here was whether
this interaction was attributable to differences within form between the two examination occasions. Examination of the cell (form by administration order) means and application of Scheffe's multiple comparison procedure revealed that these differences were not significant (p<.05). The significant interaction was attributable to differences between means on the two forms within schools in which the two test forms were administered in the same order. In the sample of eight schools in which Form B was administered first, the Form B means exceeded the Form A means in all but one school (a small school: n=9). In the 3 schools in which Form A was administered first, the Form B means were greater (see Table 7).

Table 7
Cell Means

<table>
<thead>
<tr>
<th></th>
<th>AB</th>
<th>BA</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>84</td>
<td>202</td>
</tr>
<tr>
<td>Form means</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Form A</td>
<td>29.12 (56.00%)</td>
<td>28.51 (54.83%)</td>
</tr>
<tr>
<td>Form B</td>
<td>31.46 (60.50%)</td>
<td>29.42 (56.58%)</td>
</tr>
</tbody>
</table>

Deletion of this school from the analysis revealed little change in the results. Consequently, given the lack of known cause for this one school's results and the lack of significant change when this school was removed from the sample, the students in this school were retained.
CHAPTER 5

ANALYSES AND RESULTS

Chapter Five begins with the presentation of the psychometric characteristics of the two forms. The characteristics of the two forms are then used to justify the equating techniques used. The second portion of the chapter consists of a description of the analyses and presentation of results in the sequence in which the analyses were conducted. The chapter concludes with an exploration of the source of the nonequivalence identified.

Pre-Equating Analyses

Statistical Properties of the Forms

Presented in Table 8 are the mean, standard deviation, range, internal consistency (Hoyt, 1941), standard error of measurement (SEM), skewness, and kurtosis of the distributions of scores on Form A and Form B. As shown, the mean scores of Form A, 28.69 (55.17%), and Form B, 30.02 (57.73%), are significantly different (t=5.02, p<.05), while the standard deviations (t=.58, p<.56) and internal consistency estimates (t=0.92, p<.36) are not (see Chapter 1, Hypotheses 1, 2, and 3). The value of skewness and kurtosis for Form A and for Form B are, respectively, within 1 standard error of each other. The intercorrelation coefficient corrected for attenuation (.97) exceeds the value of .95 that Linn (1975) describes as a commonly accepted cut off for equating studies (see Chapter 1, Hypothesis 4).
Table 8
Psychometric Properties of the Two Forms
n=286

<table>
<thead>
<tr>
<th></th>
<th>Form A</th>
<th>Form B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>28.69</td>
<td>30.02</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>9.46</td>
<td>8.97</td>
</tr>
<tr>
<td>Minimum score</td>
<td>9.00</td>
<td>12.00</td>
</tr>
<tr>
<td>Maximum score</td>
<td>51.00</td>
<td>51.00</td>
</tr>
<tr>
<td>Internal consistency</td>
<td>.89</td>
<td>.87</td>
</tr>
<tr>
<td>SEM</td>
<td>3.14</td>
<td>3.15</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.24</td>
<td>0.28</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2.13</td>
<td>2.24</td>
</tr>
<tr>
<td>Correlation coefficient</td>
<td>.85</td>
<td></td>
</tr>
<tr>
<td>Disattenuated correlation</td>
<td>.97</td>
<td></td>
</tr>
</tbody>
</table>

Equating

Linear Equating Angoff's Design II

Since the internal consistency of the two tests did not differ significantly, and the intercorrelation coefficient corrected for attenuation (.97) exceeded .95, Lord's (1950) linear equating method for equally reliable tests for Design II (Angoff, 1982; 1971) was used. Although Angoff described a variant of Design II in which the data are pooled across order of administration (see Chapter 2), the Form A and Form B samples were not initially combined. Before equating, the weighted means and variances for each form to be used for the linear equating were compared with the pooled estimates that Angoff suggested could be used. As shown in Table 9, the weighted means and variances and the corresponding pooled means and variances for Form A and for Form
B were found to be identical (less than 0.1% difference in the Form A and Form B variances). The equivalence of the weighted and pooled estimates means the equating results based on the pooled variances would be identical to those obtained using the weighted variances. All equatings were completed using the pooled data.

Table 9
Weighted and Pooled Means and Variances

<table>
<thead>
<tr>
<th>Sample</th>
<th>number of cases</th>
<th>mean</th>
<th>variance</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AB</td>
<td>84</td>
<td>29.12</td>
<td>99.60</td>
<td>9.98</td>
</tr>
<tr>
<td>BA</td>
<td>202</td>
<td>28.51</td>
<td>85.56</td>
<td>9.25</td>
</tr>
<tr>
<td>Weighted</td>
<td>286</td>
<td>28.69</td>
<td>89.42</td>
<td>9.46</td>
</tr>
<tr>
<td>Pooled</td>
<td>286</td>
<td>28.69</td>
<td>89.39</td>
<td>9.46</td>
</tr>
<tr>
<td>Form B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AB</td>
<td>84</td>
<td>31.46</td>
<td>88.74</td>
<td>9.42</td>
</tr>
<tr>
<td>BA</td>
<td>202</td>
<td>29.42</td>
<td>76.04</td>
<td>8.72</td>
</tr>
<tr>
<td>Weighted</td>
<td>286</td>
<td>30.02</td>
<td>80.33</td>
<td>8.96</td>
</tr>
<tr>
<td>Pooled</td>
<td>286</td>
<td>30.02</td>
<td>80.41</td>
<td>8.97</td>
</tr>
</tbody>
</table>

Linear Equating

The results of the linear equating using the pooled estimates are summarized in Table 10 and illustrated in Figure 1. As shown in Table 10, while the means of the derived score on Form B and the obtained scores on Form B are equal, the CRMS is large, 4.87. The value for the AAD is also large, 3.74. Further, the corresponding ranges of the differences between B and B*, -19 to 13, is wide.
Table 10
Results of the Linear Equating

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score A</td>
<td>28.69</td>
<td>9.46</td>
<td>9</td>
</tr>
<tr>
<td>Score B</td>
<td>30.02</td>
<td>8.97</td>
<td>12</td>
</tr>
<tr>
<td>Score B*</td>
<td>30.01</td>
<td>8.97</td>
<td>12</td>
</tr>
<tr>
<td>B-B*</td>
<td>0.01</td>
<td>4.87</td>
<td>-19</td>
</tr>
<tr>
<td>AAD</td>
<td>3.74</td>
<td>3.11</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 1 is a graph of derived Form B scores vs Form A scores. The identity function, B=A, has been added to the graph to serve as a point of reference as this is the line that would result if the two forms were perfectly equivalent.

Examination of Figure 1 reveals that the significant difference of -1.33 (2.56%) observed between the mean on Form A and the mean on Form B (Table 10) does not appear to be constant across all performance levels. That is, the equating line is not parallel to the linear identity function B=A. In particular, students at the lower performance levels (below a score of 26 (50.0%) on Form A) would have received derived scores 2 marks (3.8%) above that had there been perfect equivalence. In contrast, for those in the middle range of performance the difference was about 1 mark while the difference between the derived score and the score received had there been perfect equivalence for the the most able examinees (above a score of 39 (75.0%) on Form A) was essentially zero.
These linear equating results lend support to that conclusion of nonequivalency derived from the significant difference of the means, and adds the suggestion that the nonequivalency arises because of the poor fit between the derived score and observed scores on Form B for the less able examinees. Collectively, these results confirm the finding that the two forms were not equivalent for the sample of students included in this study.

Equipercentile Equating

The steps followed to complete the equipercentile equating were as follows:
1. A table of relative cumulative frequencies was prepared separately for the distribution of scores on Form A and the distribution of scores on Form B.

2. The cumulative frequencies and the raw scores for each distribution were plotted on arithmetic graph paper with raw scores placed on the horizontal axis. Hand smoothed curves were drawn through the plots of the two distributions.

3. Score values from the smoothed plots were read and recorded for each percentile within each distribution.

4. The score values from step 3 for Form A were plotted against the score values for Form B and a smoothed curve drawn.

5. A table of equivalent score values was prepared from this final curve. (Angoff, 1971, pp. 571-576)

Tables of cumulative frequencies and score values (step 1, step 3, and step 5) are provided in Appendix F.

Equipercentile equating results are summarized in Table 11 and illustrated in Figure 2. The means for Form A and Form B and the derived scores on B along with their corresponding standard deviations are reported in Table 11 together with the values of the CRMS and AAD.

The mean difference, B-B*, (0.18) is likely due to rounding error and smoothing error at the various stages of the equipercentile equating process. The CRMS is large, 4.86. The value for the AAD is also large, 3.73. Further, the corresponding ranges of the differences between B and B*, -20 to 13, is wide.
Table 11
Results of Equipercentile Equating

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score A</td>
<td>28.69</td>
<td>9.46</td>
<td>9</td>
</tr>
<tr>
<td>Score B</td>
<td>30.02</td>
<td>8.97</td>
<td>12</td>
</tr>
<tr>
<td>Score B*</td>
<td>29.84</td>
<td>8.91</td>
<td>11</td>
</tr>
<tr>
<td>B-B*</td>
<td>0.18</td>
<td>4.86</td>
<td>-20</td>
</tr>
<tr>
<td>AAD</td>
<td>3.73</td>
<td>3.10</td>
<td>0</td>
</tr>
</tbody>
</table>

Examination of Figure 2 reveals that the significant difference of -1.33 (2.56%) observed between the mean on Form A and the mean on Form B (Table 11) does not appear to be constant across all performance levels. That is, the equating line is not parallel to the linear identity function B=A.
In particular, students at the lower performance levels (below a score of 22 (42.3%) on Form A) would have received derived scores 2 marks (3.8%) above that had there been perfect equivalence. In contrast, for those in the middle range of performance the difference was about 1 mark, while the difference between the derived score and the score received had there been perfect equivalence for the the most able examinees (above a score of 37 (71.2%) on Form A) was essentially zero. These equipercentile equating results lend support to that conclusion of nonequivalency derived from the significant difference of the means and adds the suggestion that the nonequivalency arises because of the poor fit between the derived score and observed scores on Form B for the less able examinees.

Comparison of the Two Equatings

The results of the linear and equipercentile equatings were very similar. The conditional root mean squares for the two equatings were 4.87 (9.37%), and 4.86 (9.35%) respectively. In each case a mean absolute difference between observed score on Form B and predicted score on Form B were 3.74 (7.19%) and 3.73 (7.17%) respectively.

The two equating results also suggest the two forms perform differently for varying ability levels. There is approximately a 2 mark difference at the lower level of examination performance, about a 1 mark difference for the middle range of scores, and no difference for the upper level of achievement.
Earlier in the chapter, given that the linear weighted and linear pooled means and variances agreed, the decision to pool was taken. Using the pooled data, the equipercentile equating results agreed with the results of the linear equating. As there is no advantage to equipercentile equating as demonstrated by the CRMS and the AAD, all subsequent analyses were conducted using linear equating with the pooled sample of 286 students.

**Differential Performance at Lower Achievement Levels**

Both equatings revealed greater differences in the performance on the two forms by lower achieving students. This is contrary to a suggestion by Petersen, Kolen, and Hoover (1989) that test forms may appear equivalent at chance levels of achievement (p. 243). The interpretability of scores, particularly the lower scores, may be questionable due to the phenomenon of guessing. If these lower scores are suspect, then the significant differences found may be in fact due to guessing. Various studies indicate, however, there is no clear advantage to either applying a correction for chance formula or removing chance scores from the sample (Angoff, 1989; Albanese, 1988; Donlon, 1981; Tinkelman, 1971).

In contrast, the performance of high scoring students was similar on both forms. This may be explained by the possible presence of a ceiling effect. If both forms are too easy, then high performing students would be expected to achieve similar and near perfect scores. However, the observed score distributions do not show the strong negative skewing (see Table 8) that would be
expected if a ceiling effect was present. Therefore this explanation was rejected.

Assessment of Equivalence of Forms

CRMS and AAD

The conditional root mean square, CRMS, was described in Chapter 2 as an indicator of the degree to which a score earned by a student would differ from a score derived for that student from an equating table. Similarly the absolute average difference, AAD, provides an indication of the degree to which an individual would achieve the same score on the two forms to be equated.

As reported in Table 8, the standard error of measurement (SEM) was 3.15 (6.1%) and the CRMS was 4.87 (9.37%) (Table 10). The SEM provides a logical lower bound for the CRMS. Form B cannot be more parallel to Form A than Form A is to itself, nor can Form A be more parallel to Form B than Form B is to itself. The CRMS was 1.55 times greater than the SEM.

The AAD, like the CRMS, can serve as an indicator of usefulness or interpretability of the parallel form scores. As reported in Table 10, the AAD was 3.73 (7.17%). In the ideal situation, the AAD would be zero. This value is judged too large to be an indication that Form A and Form B are parallel forms.

Standard Error of Equating

The standard error of equating ($s_{b*}$) reflects the degree to which the equating results would vary if the equating was repeated with a different sample. The modified Dorans and Lawrence (1990) approach described in Chapter 2 allows a
graphical interpretation of equivalence. As stated there, the confidence bands are established about a line that would be obtained given exact equivalency. As shown in Figure 3, the plot of the equating function cuts through the 95% confidence band limits about the identity function. As the difference between the equating function and the identity function exceeds the standard error of equating, the two forms are judged to be overall nonequivalent (see Chapter 1, Hypotheses 5).

Figure 3 Standard Error of Equating
Total Examination

As suggested by Figures 1 and 2, and confirmed by Figure 3, while the forms are equivalent for the highest achievers, for most of the
score range, a difference that is too large to be attributed to equating error is found.

Nonequivalence

Source of Nonequivalence

As described in Chapter 3, prior p value information was available and used to form 24 pairs of items. One item from each pair was randomly assigned to each form. An additional 17 pairs of items in which only one item had a prior p value were created. Again, one item from each pair was assigned to each form so that 8 items with prior p value information were placed on one form and 9 items on the other form. For the remaining 11 pairs of items judgements were made about item p value similarity prior to item assignment to form. The three different conditions of p value information within item pairs were the basis for the formation of three subtests.

Equating of Subtests

It was hypothesized that the lack of parallelism noted in the two forms was attributable to differences arising from an inability to match items on difficulty level when p values were not available. To test this hypothesis, the linear equating was replicated for each of the three subtests now formed within each test. The equating results were examined using the modified Dorans and Lawrence (1990) graphical analysis. As the separate subtests are comprised of varying numbers of items, all reported figures on the graphs are percentages in order to facilitate comparison among the three.
As previously described, the first pair of subtests each consisted of 24 items. Difficulty indices (p values) were known for all these items prior to administration. The mean p value for Form A subtest items was .64 while for the Form B subtest the mean p value was .65. Equivalent subtests were expected. As shown in Figure 4, the equating function lies between the standard error of equating bands, ±1.96s_b*, the two forms comprised of pairs of items for which p values were known for both items are equivalent.

Figure 4 Standard Error of Equating
P Values Known for Both Items
The second set of subtests consisted of 17 items each. Eight items with known p values were placed on Form A while the remaining nine items with known p values were placed on Form B.

The Form A subtest mean p value was .62 for the items with prior p values. The Form B subtest had a mean p value of .63 for the items with prior p values. As with the first pair of subtests equivalent tests were expected. As seen in Figure 5, subtests in which the p value for only one of the pair of items were known also resulted in equivalent forms.

Figure 5 Standard Error of Equating
P Value Known One Item Only

The third set of subtests consisted of 11 items each. None of these items had prior p values. The results for these subtests formed from pairs of items for which the difficulty was
qualitively judged are markedly different (see Figure 6). These subtests are not equivalent. The source of nonequivalence of the total forms appears to be due to the pairs of items for which p value data was not available for either item.

Figure 6 Standard Error of Equating
P Values Unknown for Both Items
CHAPTER 6
CONCLUSIONS

This chapter contains four sections. First, a summary of the study including its purpose, procedure, and findings is given. Limitations of the study are then discussed. The limitations are followed by the conclusions of the study. The chapter ends with implications for practice and implications for research.

Summary

Purpose and Problem

Questions have been raised concerning the equivalency of the January, June, and August forms of the British Columbia provincial Grade 12 examinations for a given subject. The implementation of a new procedure for constructing these forms began in the 1990/91 school year. The change in procedure was expected to improve the ability of the examination construction team to produce equivalent forms. The purpose of this study was to duplicate this new procedure and assess the equivalency of the forms that result.

Procedure

An examination construction team, all of whom had previous experience with the British Columbia Ministry of Education's Student Assessment Branch, simultaneously constructed two forms of a Biology 12 examination from a common table of specifications using a pool of multiple choice items from previous examinations. Some of the items had accompanying p value information at the time of examination construction, while other
items lacked this information. A sample of students was obtained in the Okanagan, Thompson, and North Thompson areas of British Columbia. Both forms were administered to each student, as required by the test equating design (Design II (Angoff, 1971)) chosen. The final usable data sample consisted of responses from 286 students, with Form A administered first to a subsample of 84 students and Form B administered first to the other 202 students.

Analysis and Results

The data were first analyzed using a classical item analysis followed by a $2\times2$ order-by-form fixed effects ANOVA with repeated measures on the second factor. Classical item analysis revealed all items on both forms performed satisfactorily, ruling out the alternate hypothesis of a flawed item(s) being the cause of the lack of equivalence found. The ANOVA revealed a significant difference in the means of the two forms. While no significant order effect was found, a significant order-by-form interaction was present. No other significant differences between the two forms were found.

Linear and equipercentile equatings were carried out. Since the weighted and pooled estimates for the variances of the two forms were found to be equal, the pooled estimates were used for these analyses. Both equatings yielded essentially the same results with the same degree of agreement between observed scores and derived scores (B derived from A) in the sample as estimated by the CRMS and the AAD. Both revealed that students at high levels of achievement found the forms of equal difficulty
while students at lower levels of achievement found one form to be easier than the other form. Since the linear and equipercentile method yielded similar results, and in light of the arbitrariness associated with equipercentile equating, all further analyses were conducted using the linear equating method and computation of standard error of equating. The forms were judged to be nonequivalent as the CRMS and AAD were large and a plot of the equating function deviated around the identity function fell outside the range of the standard error of equating estimated for the sample.

The source of the nonequivalency was examined by separating each of the two forms into three subtests based on the items possessing or lacking p values at the time of test construction. The first subtest pair of forms consisted of 24 pairs of items with prior p value information available for both items in the pair. The second subtest pair of forms were created from an additional 17 pairs of items in which only one item of the pair had a prior p value. The third pair of subtest forms were created from the remaining 11 pairs of items for which qualitative judgements had been made about item difficulty prior to item assignment to form. The forms were judged to be equivalent or nonequivalent based on whether the plot of the equating function deviated around the identity function fell inside or outside the range of standard error of equating estimated for the sample. The first pair of subtests were judged to be equivalent, as were the second pair of subtests.
Only the pair of subtests formed from items with no p value present for either item were judged to be nonequivalent.

Limitations

The generalizability of the study is limited by several factors. This study was restricted to multiple choice items in the core curriculum of one subject area. Factors not considered were supply items, optional areas within a subject area, and other subject areas. However, since the multiple choice items on the core areas are most probably the most reliable and stable items, if two equivalent forms of a test using only these items cannot be constructed, there is little hope that equivalency using the other forms of items or other topic areas can be achieved.

Regarding the sample, although there was sufficient range in scores to examine the student performance at all plausible achievement levels (scores ranged from 17% to 98%) for the experimental examinations, the score distributions obtained cannot be assumed to represent provincial norms.

Conclusions

The central problem of this study was: will the new procedure for the construction of equivalent forms from a common table of specifications result in equivalent forms? The major conclusion is that the procedure in its present state cannot be relied on to produce equivalent forms. Subsequent investigation following the finding of nonequivalent forms suggests that the lack of equivalency results from the inability of an experienced examination assembly team to accurately match levels of
difficulty for pairs of items which do not have prior item statistics accompanying them. Conversely, an experienced examination assembly team is able to produce equivalent forms if prior p value information is available. This study utilized an item pool in which the proportion of the items with prior p value information was greatly in excess of the 20% goal for the 1990 examinations (see Appendix A). The lack of equivalency would be expected to worsen when the proportion of items with prior p values is decreased.

Equivalency of forms can be judged using a combination of a fit statistic, the conditional root mean square (CRMS), and a measure of sample error in the equating, the standard error of equating ($s_y^*$).

Implications

Implications for Practice

The procedure for examination construction used in this study cannot be expected to produce equivalent forms within a given subject area. The use of a trained and experienced examination construction team will not guarantee equivalent forms, items with p value information must be used.

The magnitude of the CRMS and AAD indicate that a student's score on one form cannot be taken as a reasonable estimate of the score that would have been obtained had a different form been written. This is of particular importance when interpreting the student scores obtained on the August supplemental examination. The score from the August supplemental examination is used in
place of a January or June score that is either absent or unsatisfactory. The group writing the August supplemental cannot be assumed to be randomly equivalent to the larger population.

The awarding of provincial scholarships depend in part on scores achieved on these examinations. Scores from students writing an examination in January, for example Biology 12, must be compared with scores for another student writing a different Biology 12 examination in June. The modified Dorans and Lawrence (1990) procedure could be used to judge January and June forms as equivalent or nonequivalent in the score range of interest. If a pair of forms are equivalent, no equated score need be used, the scores are interchangeable. If the January and June forms are judged as nonequivalent in the score range of interest, then an equated score could be given to the January student. In either situation, the January and June examinees in a given subject area can be compared.

Implications for Future Research

The implication that the use of classical item statistics could be used to produce parallel forms should be tested for multiple mark (supply) questions. The possibility of creating statistically parallel optional sections within an examination form should also be explored. Other subject areas should be investigated. There should be an attempt to duplicate the results that suggested equivalent examination forms were produced when item pairs in which only one member of a pair possessed prior item statistics were used.
REFERENCES


Appendix A

Ministry Examination Construction Process
FLOWCHART and TIMELINES
for the
Production of the 1991 Examinations

Phase 1
EXAM WRITING Team develop
items sufficient for
THREE Prov. & THREE Schol. exams
(mtg. req'd. for TEAM EDIT.),
4 persons @ $2000/person
Staggered submission:
1st-April 30 1990
2nd-July 15 or before.

DESKTOP PUBLISHING
Word process input submitted by Exam Writing teams.

POOL of ITEMS
for
PREPARATION TEAM,
sufficient to produce
THREE SETS of exams,
(set = Prov.&Schol.)

SELECTION of items to prepare
THREE SETS of EXAMS
by the
EXAM PREPARATION TEAM,
4 persons @ $3150/exam/person
If possible 20% of items on each exam
should be "item-bank field tested"
Team meeting sometime in August.

ITEMS (field-tested?)
from:
ITEM-BANKING
20% of items on each exam should be
"item-bank field tested"
The potential of these items as yet is
undetermined. For maximum flexibility, Exam
Developers not depend on this as a source;
rather these items should appraised as they
become available and used where
appropriate. The 20% mentioned above
represents the ideal goal.

DESKTOP PUBLISHING UNIT
of the 3 sets (Prov.&Schol)
Into Review-ready first drafts.
EXTERNAL REVIEW of the 3 sets of exams; 3 day mg, req'd. Sat.&Sun. Mon. or Fri. 4 persons @ $120 per set of exams (Prov. Schol) 90/10/15

Suggested changes from External Review to PREP TEAM for possible action (changes).

DESKTOP PUBLISHING UNIT
Word process Review crte. changes to produce second draft for external review.

JOINT REVIEW of the 3 sets of exams; 3 day mg, req'd. Sat.&Sun. Mon. or Fri. consisting of 4 Reviewers and 4 Preparers.

DESKTOP PUBLISHING UNIT
Word process 'Review team changes' to produce Internal Review draft.

INTERNAL REVIEW SAB
N.B. The DIPU will also be involved processing changes as they arise at this time. Because of time, only 1 set of exams can be processed with these deadlines, perhaps the other 2 should be staggered for the two weeks following.

TRIAL WRITING
of 3 sets of exams ($60 per exam)

QUEEN'S PRINTER
Camera-ready copies of ONE (these) set(s) of exams sent to OP for Blue-line production.
Appendix B

Materials for Construction Team
# TABLE OF SPECIFICATIONS

<table>
<thead>
<tr>
<th>TOPIC</th>
<th>CURRICULUM GUIDE REFERENCE</th>
<th>COGNITIVE LEVEL</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Knowledge</td>
<td></td>
</tr>
<tr>
<td>Methods</td>
<td></td>
<td>Understanding</td>
<td></td>
</tr>
<tr>
<td>1. Experimental Design</td>
<td></td>
<td>Higher Mental Processes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TOTAL</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Q</td>
<td>10</td>
</tr>
<tr>
<td>2. Cells</td>
<td>III Cell Compounds</td>
<td>45</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>IV Ultrastructure</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>V Ultraprocesses</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VI Photosynthesis</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>19%</td>
<td>14</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TOTAL</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(24%)</td>
<td>(56%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100%</td>
<td>82%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>48%</td>
<td>written response questions</td>
</tr>
</tbody>
</table>

*METHODS AND PRINCIPLES.* Questions related to Experimental Design and Homeostasis will be embedded within topics 2 and 3 in the above table, and account for 12.22% of the content of the exam. As a result, horizontal addition of topics 2 and 3 yield 24% and 51% respectively.
### TABLE 5A
### MULTIPLE-CHOICE SUMMARY
### JANUARY 1987  N = 1,924

<table>
<thead>
<tr>
<th>Item No.</th>
<th>Topic and Cog. Level</th>
<th>% of Students Choosing Alternatives(^1)</th>
<th>Item No.</th>
<th>Topic and Cog. Level</th>
<th>% of Students Choosing Alternatives(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>1</td>
<td>1K</td>
<td>73(^*)</td>
<td>4</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>2</td>
<td>1K</td>
<td>16</td>
<td>11</td>
<td>21</td>
<td>53(^*)</td>
</tr>
<tr>
<td>3</td>
<td>3U</td>
<td>24</td>
<td>72(^*)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>1K</td>
<td>4</td>
<td>88(^*)</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>1K</td>
<td>0</td>
<td>3</td>
<td>74(^*)</td>
<td>22</td>
</tr>
<tr>
<td>6</td>
<td>1H</td>
<td>46(^*)</td>
<td>33</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>2U</td>
<td>12</td>
<td>4</td>
<td>82(^*)</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>2U</td>
<td>9</td>
<td>4</td>
<td>68(^*)</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>1K</td>
<td>40(^*)</td>
<td>17</td>
<td>20</td>
<td>22</td>
</tr>
<tr>
<td>10</td>
<td>3K</td>
<td>4</td>
<td>34</td>
<td>48(^*)</td>
<td>13</td>
</tr>
<tr>
<td>11</td>
<td>3K</td>
<td>14</td>
<td>4</td>
<td>12</td>
<td>70(^*)</td>
</tr>
<tr>
<td>12</td>
<td>5U</td>
<td>24</td>
<td>16</td>
<td>52(^*)</td>
<td>8</td>
</tr>
<tr>
<td>13</td>
<td>5K</td>
<td>7</td>
<td>3</td>
<td>82(^*)</td>
<td>9</td>
</tr>
<tr>
<td>14</td>
<td>5K</td>
<td>41</td>
<td>45(^*)</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>15</td>
<td>5U</td>
<td>29</td>
<td>12</td>
<td>50(^*)</td>
<td>9</td>
</tr>
<tr>
<td>16(^1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>4X</td>
<td>41</td>
<td>6</td>
<td>15</td>
<td>38(^*)</td>
</tr>
<tr>
<td>18</td>
<td>4X</td>
<td>43(^*)</td>
<td>23</td>
<td>9</td>
<td>24</td>
</tr>
<tr>
<td>19</td>
<td>4U</td>
<td>62(^*)</td>
<td>14</td>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>7K</td>
<td>5</td>
<td>71(^*)</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>21</td>
<td>7U</td>
<td>12</td>
<td>74(^*)</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>

See p. 6 for 43-70 (Optional Section)

\(^*\) correct response (The % students choosing the correct response indicates the difficulty level of the question.)

\(^1\) These codes represent topics and cognitive levels listed in Table 4. For further discussion of topics and cognitive levels, see Page 1.

\(^2\) Percentages are based on all students who responded to the item.

\(^3\) Item 16 was deleted at the recommendation of the markers.
### Table 53
**Multiple-Choice Summary**
**June 1987 N = 7,288**

<table>
<thead>
<tr>
<th>Item No.</th>
<th>Topic and Cog. Level</th>
<th>% of Students Choosing Alternatives²</th>
<th>Item No.</th>
<th>Topic and Cog. Level</th>
<th>% of Students Choosing Alternatives²</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1K</td>
<td>A: 2 B: 9 C: 3 D: 86*</td>
<td>22</td>
<td>7U</td>
<td>A: 1 B: 78* C: 7 D: 14</td>
</tr>
<tr>
<td>2</td>
<td>2K</td>
<td>A: 17 B: 19 C: 19 D: 45*</td>
<td>23</td>
<td>7U</td>
<td>A: 68* B: 11 C: 4 D: 16</td>
</tr>
<tr>
<td>3</td>
<td>1U</td>
<td>A: 80* B: 7 C: 11 D: 2</td>
<td>24</td>
<td>7K</td>
<td>A: 68* B: 15 C: 14 D: 3</td>
</tr>
<tr>
<td>4</td>
<td>1K</td>
<td>A: 8 B: 2 C: 87* D: 3</td>
<td>25</td>
<td>7U</td>
<td>A: 6 C: 11 D: 51* D: 32</td>
</tr>
<tr>
<td>5</td>
<td>3U</td>
<td>A: 48* B: 6 C: 2 D: 44</td>
<td>26</td>
<td>7U</td>
<td>A: 29* B: 21 C: 26 D: 24</td>
</tr>
<tr>
<td>6</td>
<td>3H</td>
<td>A: 5 B: 22 C: 16 D: 57*</td>
<td>27</td>
<td>7K</td>
<td>A: 3 B: 3 C: 55* D: 39</td>
</tr>
<tr>
<td>8</td>
<td>3U</td>
<td>A: 63* B: 3 C: 23 D: 11</td>
<td>29</td>
<td>8U</td>
<td>A: 13 B: 51* C: 28 D: 8</td>
</tr>
<tr>
<td>9</td>
<td>3U</td>
<td>A: 3 B: 3 C: 86* D: 8</td>
<td>30</td>
<td>8U</td>
<td>A: 10 B: 68* C: 18 D: 4</td>
</tr>
<tr>
<td>10</td>
<td>3U</td>
<td>A: 10 B: 21 C: 6 D: 65*</td>
<td>31</td>
<td>8U</td>
<td>A: 48* B: 5 C: 19 D: 27</td>
</tr>
<tr>
<td>11</td>
<td>3U</td>
<td>A: 13 B: 71* C: 8 D: 8</td>
<td>32</td>
<td>8K</td>
<td>A: 17 B: 78* C: 1 D: 4</td>
</tr>
<tr>
<td>12</td>
<td>5K</td>
<td>A: 73* B: 18 C: 7 D: 2</td>
<td>33</td>
<td>8K</td>
<td>A: 17 B: 19 C: 22 D: 42*</td>
</tr>
<tr>
<td>13</td>
<td>5K</td>
<td>A: 15 B: 35 C: 43* D: 7</td>
<td>34</td>
<td>8U</td>
<td>A: 6 B: 25 D: 65* D: 4</td>
</tr>
<tr>
<td>14</td>
<td>4U</td>
<td>A: 21 B: 3 C: 6 D: 70*</td>
<td>35</td>
<td>9U</td>
<td>A: 1 B: 8 D: 71* D: 19</td>
</tr>
<tr>
<td>16</td>
<td>5K</td>
<td>A: 9 B: 7 D: 74* D: 11</td>
<td>37</td>
<td>9U</td>
<td>A: 20 B: 47* C: 10 D: 23</td>
</tr>
<tr>
<td>17</td>
<td>5K</td>
<td>A: 33 B: 16 C: 14 D: 37*</td>
<td>38</td>
<td>10U</td>
<td>A: 3 B: 77* C: 18 D: 2</td>
</tr>
<tr>
<td>18*</td>
<td>5U</td>
<td>A: 1 B: 16 C: 4 D: 78*</td>
<td>39</td>
<td>10U</td>
<td>A: 2 B: 86* C: 2 D: 10</td>
</tr>
<tr>
<td>19</td>
<td>5U</td>
<td>A: 9 B: 75* C: 10 D: 6</td>
<td>40</td>
<td>10K</td>
<td>A: 6 B: 89* C: 4 D: 1</td>
</tr>
<tr>
<td>20</td>
<td>7K</td>
<td>A: 6 B: 8 D: 54* D: 32</td>
<td>41</td>
<td>10U</td>
<td>A: 71* B: 11 D: 9 D: 9</td>
</tr>
</tbody>
</table>

See p. 6 for 43-70 (Optional Section)

*Correct response (The % students choosing the correct response indicates the difficulty level of the question.)*

¹These codes represent topics and cognitive levels listed in Table 4. For further discussion of topics and cognitive levels, see Page 1.

²Percentages are based on all students who responded to the item.

³Item 18 was deleted at the recommendation of the markers.
**TABLE 5A**

MULTIPLE-CHOICE SUMMARY

JANUARY 1988  N = 2,139

<table>
<thead>
<tr>
<th>Item No.</th>
<th>Topic and Cog. Level</th>
<th>% of Students Choosing Alternatives</th>
<th>Item No.</th>
<th>Topic and Cog. Level</th>
<th>% of Students Choosing Alternatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2U</td>
<td>64% 27 4 5</td>
<td>27</td>
<td>6X</td>
<td>9 68% 19 5</td>
</tr>
<tr>
<td>2</td>
<td>1U</td>
<td>10 45% 39 6</td>
<td>28</td>
<td>6X</td>
<td>11 19 28% 42</td>
</tr>
<tr>
<td>3</td>
<td>1K</td>
<td>8 3 10 79%</td>
<td>29</td>
<td>6X</td>
<td>36% 38 16 10</td>
</tr>
<tr>
<td>4</td>
<td>1U</td>
<td>7 3 84% 5</td>
<td>30</td>
<td>6U</td>
<td>7 59% 4 29</td>
</tr>
<tr>
<td>5</td>
<td>1H</td>
<td>61% 10 21 9</td>
<td>31</td>
<td>6U</td>
<td>12 22 7 59%</td>
</tr>
<tr>
<td>6</td>
<td>1U</td>
<td>11 65% 14 10</td>
<td>32</td>
<td>7U</td>
<td>35 43% 11 9</td>
</tr>
<tr>
<td>7</td>
<td>1K</td>
<td>77% 10 10 3</td>
<td>33</td>
<td>7U</td>
<td>4 2 92% 3</td>
</tr>
<tr>
<td>8</td>
<td>2U</td>
<td>68% 4 3 25</td>
<td>34</td>
<td>7U</td>
<td>5 0 6 89%</td>
</tr>
<tr>
<td>9</td>
<td>2U</td>
<td>2 65% 20 12</td>
<td>35</td>
<td>7U</td>
<td>6 23 18 53%</td>
</tr>
<tr>
<td>10</td>
<td>3H</td>
<td>42% 49 3 6</td>
<td>36</td>
<td>7U</td>
<td>83% 4 11 3</td>
</tr>
<tr>
<td>11</td>
<td>3K</td>
<td>15 65% 7 13</td>
<td>37</td>
<td>7K</td>
<td>13 10 58% 19</td>
</tr>
<tr>
<td>12</td>
<td>3U</td>
<td>45% 14 21 21</td>
<td>38</td>
<td>8K</td>
<td>3 87% 3 7</td>
</tr>
<tr>
<td>13</td>
<td>3K</td>
<td>3 2 79% 16</td>
<td>39</td>
<td>8K</td>
<td>3 6 81% 9</td>
</tr>
<tr>
<td>14</td>
<td>3U</td>
<td>12 4 15 70%</td>
<td>40</td>
<td>10U</td>
<td>32 7 50% 12</td>
</tr>
<tr>
<td>15</td>
<td>3U</td>
<td>5 14 13 68%</td>
<td>41</td>
<td>10K</td>
<td>20 1 1 78%</td>
</tr>
<tr>
<td>16</td>
<td>4U</td>
<td>10 70% 8 12</td>
<td>42</td>
<td>10U</td>
<td>5 77% 5 13</td>
</tr>
<tr>
<td>17</td>
<td>4U</td>
<td>13 45% 11 31</td>
<td>43</td>
<td>10U</td>
<td>9 5 70% 17</td>
</tr>
<tr>
<td>18</td>
<td>4U</td>
<td>69% 24 4 3</td>
<td>44</td>
<td>10K</td>
<td>54% 21 17 8</td>
</tr>
<tr>
<td>19</td>
<td>4U</td>
<td>6 59% 24 10</td>
<td>45</td>
<td>10K</td>
<td>17 11 21 51%</td>
</tr>
<tr>
<td>20</td>
<td>4K</td>
<td>19 8 17 56%</td>
<td>46</td>
<td>10K</td>
<td>1 10 88% 1</td>
</tr>
<tr>
<td>21</td>
<td>4U</td>
<td>9 77% 2 11</td>
<td>47</td>
<td>10U</td>
<td>59% 21 6 14</td>
</tr>
<tr>
<td>22</td>
<td>5K</td>
<td>19 11 50% 20</td>
<td>48</td>
<td>9U</td>
<td>2 12 67% 19</td>
</tr>
<tr>
<td>23</td>
<td>5U</td>
<td>14 4 79% 2</td>
<td>49</td>
<td>9K</td>
<td>9 12 51% 29</td>
</tr>
<tr>
<td>24</td>
<td>5K</td>
<td>39% 14 10 39</td>
<td>50</td>
<td>9U</td>
<td>22 51% 18 9</td>
</tr>
<tr>
<td>25</td>
<td>5U</td>
<td>26 4 69% 1</td>
<td>51</td>
<td>9U</td>
<td>72% 19 3 6</td>
</tr>
<tr>
<td>26</td>
<td>5K</td>
<td>43% 42 7 8</td>
<td>52</td>
<td>9U</td>
<td>2 13 16 68%</td>
</tr>
</tbody>
</table>

*Correct response (The % of students choosing the correct response indicates the difficulty level of the question).

1These codes represent topics and cognitive levels listed in Table 4.

2Percentages are based on all students who responded to the item.
### TABLE 5B
MULTIPLE-CHOICE SUMMARY
JUNE 1988 N = 7,992

<table>
<thead>
<tr>
<th>Item No.</th>
<th>Topic and Cog. Level</th>
<th>% of Students Choosing Alternatives$^2$</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2U</td>
<td>3% 3% 84% 10%</td>
<td>3</td>
<td>3</td>
<td>84</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1H</td>
<td>53% 34% 3% 10%</td>
<td>53</td>
<td>34</td>
<td>3</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1U</td>
<td>8% 10% 72% 10%</td>
<td>8</td>
<td>10</td>
<td>72</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1U</td>
<td>7% 61% 31% 1%</td>
<td>7</td>
<td>61</td>
<td>31</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1K</td>
<td>8% 9% 11% 72%</td>
<td>8</td>
<td>9</td>
<td>11</td>
<td>72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1U</td>
<td>2% 22% 75% 1%</td>
<td>2</td>
<td>22</td>
<td>75</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1H</td>
<td>3% 11% 81% 4%</td>
<td>3</td>
<td>11</td>
<td>81</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2H</td>
<td>80% 7% 4% 8%</td>
<td>80</td>
<td>7</td>
<td>4</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>2K</td>
<td>5% 41% 41% 13%</td>
<td>5</td>
<td>41</td>
<td>41</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>3U</td>
<td>3% 27% 8% 63%</td>
<td>3</td>
<td>27</td>
<td>8</td>
<td>63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>3U</td>
<td>86% 10% 2% 3%</td>
<td>86</td>
<td>10</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>3U</td>
<td>8% 64% 4% 24%</td>
<td>8</td>
<td>64</td>
<td>4</td>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>3U</td>
<td>57% 19% 13% 11%</td>
<td>57</td>
<td>19</td>
<td>13</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>3K</td>
<td>67% 4% 18% 11%</td>
<td>67</td>
<td>4</td>
<td>18</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>5U</td>
<td>63% 11% 15% 10%</td>
<td>63</td>
<td>11</td>
<td>15</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>5U</td>
<td>12% 6% 47% 33%</td>
<td>12</td>
<td>6</td>
<td>47</td>
<td>33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>5K</td>
<td>3% 69% 18% 10%</td>
<td>3</td>
<td>69</td>
<td>18</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>5K</td>
<td>15% 9% 6% 70%</td>
<td>15</td>
<td>9</td>
<td>6</td>
<td>70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>5K</td>
<td>6% 75% 8% 11%</td>
<td>6</td>
<td>75</td>
<td>8</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>5U</td>
<td>9% 21% 16% 53%</td>
<td>9</td>
<td>21</td>
<td>16</td>
<td>53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>5U</td>
<td>37% 10% 22% 31%</td>
<td>37</td>
<td>10</td>
<td>22</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>4K</td>
<td>8% 10% 21% 61%</td>
<td>8</td>
<td>10</td>
<td>21</td>
<td>61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>5U</td>
<td>64% 27% 4% 5%</td>
<td>64</td>
<td>27</td>
<td>4</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>5U</td>
<td>4% 9% 83% 8%</td>
<td>4</td>
<td>9</td>
<td>83</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>5U</td>
<td>7% 18% 58% 16%</td>
<td>7</td>
<td>18</td>
<td>58</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>6K</td>
<td>18% 68% 13% 1%</td>
<td>18</td>
<td>68</td>
<td>13</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$correct response (The % of students choosing the correct response indicates the difficulty level of the question).

$^1$These codes represent topics and cognitive levels listed in Table 4.

$^2$Percentages are based on all students who responded to the item.
Appendix C

Examination Forms
FORM I

NAME: ___________________________ NUMBER: ____________

SCHOOL: __________________________

BIOLOGY 12

GENERAL INSTRUCTIONS

1. Write your name and school in the space provided above.

2. DO NOT WRITE YOUR NAME ON THE ANSWER SHEET.

3. Ensure that the Form Number (I or II) on this booklet and the answer sheet are the same.

4. Ensure that the I.D. Number (1 to 600) on answer sheet is the same as on the exam booklet.

5. When writing the second form (I or II) ensure that the I.D. Number is the SAME I.D. as on the first form. (I or II)

6. When instructed to open this booklet, check the numbering of the pages to ensure that they are numbered in sequence from page 1 to the last page which is identified by END OF EXAMINATION.

7. At the end of the examination, place your answer sheet inside the front cover of this booklet and return the booklet and your answer sheet to the supervisor.

Value: 52 marks (one mark per question)    Time: 40 minutes

INSTRUCTIONS:

For each question, select the BEST answer and record your choice on the answer sheet provided. Using an HB pencil, completely fill in the circle that has the letter corresponding to your answer.
Use the following diagram to answer question 1.

| Three-letter codons of messenger RNA and the amino acids specified by the codons. |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| AUG | Asparagine | CAC | Cysteine | GAC | Aspartic acid | UAC | Tyrosine |
| AAA | Lysoleucine | CAA | Glutamine | GAA | Glutamic acid | UAG | (Termination) |
| AAG | Tryptophan | CCG | Phenylalanine | GCC | Alanine | UCU | Serine |
| AGU | Serine | CGU | Arginine | GGU | Glicine | UGU | Cysteine |
| AGA | Arginine | CGU | Leucine | GGU | Glutamic acid | UGG | Tyrosine |
| AUC | Histidine | CUA | Leucine | GUC | Valine | UUU | Phenylalanine |
| AUA | Methionine | CUU | Leucine | GUC | Glutamine | UUA | Leucine |
| AUG | Methionine | CUG | Leucine | GUG | Valine | UAG | Leucine |

1. Using the above table of codons, which amino acid sequence below is formed from the DNA strand?

A T A C G A C A A G C C

A. Tyrosine-alanine-valine-arginine.
B. Tyrosine-arginine-leucine-alanine.
C. Methionine-alanine-leucine-alanine.
D. Methionine-arginine-glutamic acid-cysteine.

Use the following information to answer 2.

\[ \text{glucose + fructose} \rightarrow \text{sucrose + water} \]

2. The above reaction represents

A. reduction.
B. synthesis.
C. hydrolysis.
D. denaturation.
Use the following diagram to answer question 3.

3. The box which indicates a single nucleotide is
   A. 1.
   B. 2.
   C. 3.
   D. 4.

4. Which of the following would describe the tertiary structure of a protein?
   A. A chain of amino acids in a linear sequence.
   B. A molecule whose specific shape is spiral.
   C. A molecule characterized by side chains of amino acids.
   D. A molecule whose specific shape is determined by folding back on itself.

Use the following table to answer question 5.

<table>
<thead>
<tr>
<th></th>
<th>CELL MEMBRANE</th>
<th>CELL WALL</th>
<th>CHLOROPLAST</th>
<th>NUCLEAR MEMBRANE</th>
<th>RIBOSOME</th>
<th>MITOCHONDRION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. Which of the above structures are found in eukaryotic cells but not in prokaryotic cells?
   A. 1, 3, 5.
   B. 1, 2, 5.
   C. 2, 4, 6.
   D. 3, 4, 6.
9. The 5-carbon compound that reacts with carbon dioxide in the Calvin cycle is

A. PGA (phosphoglyceric acid).
B. RuBp (ribulose biphosphate).
C. PGAL (phosphoglyceraldehyde).
D. NADP (nicotinamide adenine dinucleotide phosphate).

Use the following diagram to answer question 10.

![Diagram showing energy levels and progress of reaction]

10. $E_a$ is greater in graph X than in graph Y. The BEST explanation for this is that the energy of activation in Y is

A. raised by the addition of an enzyme.
B. lowered by the addition of an enzyme.
C. raised by the addition of more substrate.
D. lowered by the addition of more substrate.
11. Which statement BEST explains why the graph rises at C?

A. More enzyme was added.
B. More substrate was added.
C. The temperature was lowered.
D. A competitive inhibitor was added.

12. The cell membrane surrounds some molecules in solution and takes them into the cell. This is an example of

A. diffusion.
B. pinocytosis.
C. phagocytosis.
D. facilitated transport.

13. During yeast cell fermentation, pyruvic acid is converted to

A. alcohol.
B. glycogen.
C. lactic acid.
D. active acetate.
Use the following diagram to answer question 14.

![Diagram of U-shaped tube with 4% salt solution in X and 10% salt solution in Y.](attachment:image.png)

14. If the experimental setup illustrated above were allowed to sit for 5 hours, what net movement of salt and water would occur?

A. Salt moves from Y to X; water moves from X to Y.
B. Salt moves from X to Y; water moves from Y to X.
C. Salt moves from Y to X; water moves from Y to X.
D. Salt moves from X to Y; water moves from X to Y.

Use the following diagram to answer question 15.

![Diagram of a mitochondrion showing labeled parts 1-4.](attachment:image.png)

15. In aerobic respiration, most of the ATP is produced at

A. 1.
B. 2.
C. 3.
D. 4.
16. The difference between active transport and facilitated transport is that ONLY

A. active transport requires ATP.
B. facilitated transport requires ATP.
C. active transport uses carrier molecules.
D. facilitated transport uses carrier molecules.

17. Which of the following occurs in BOTH cyclic and noncyclic photophosphorylation?

A. O₂ is produced.
B. ATP is produced.
C. CO₂ is broken down.
D. H₂O is a hydrogen donor.

Use the following diagram to answer question 18.

\[ \text{Visible light} \rightarrow \text{H}_2\text{O} \rightarrow \text{ATP} \rightarrow \text{CO}_2 \rightarrow \text{Glucose} \]

18. The name given to the pathway labelled "1" is

A. hydrolysis.
B. Krebs cycle.
C. Calvin cycle.
D. photophosphorylation.

19. The production of glucose occurs during

A. aerobic respiration.
B. anaerobic respiration.
C. the carbon dioxide-reducing reactions.
D. the light-capturing reactions.

20. Which of the following is classified as an organ?

A. Bone.
B. Skin.
C. Blood.
D. Muscle.
21. Where in the digestive tract does the chemical digestion of protein begin?
   A. Mouth.
   B. Stomach.
   C. Small intestine.
   D. Large intestine.

22. Protein is digested by
   A. lipase.
   B. trypsin.
   C. amylase.
   D. secretin.

23. Which enzyme catalyses the following reaction?
   \[ \text{fat droplets} + \text{water} \rightarrow \text{glycerol} + \text{fatty acids} \]
   A. Lipase.
   B. Pepsin.
   C. Maltase.
   D. Trypsin.

24. An example of chemical digestion is
   A. chewing.
   B. absorption.
   C. hydrolysis.
   D. peristalsis.

25. Which of the following is NOT a component of gastric juice?
   A. Water.
   B. Pepsin.
   C. Amylase.
   D. Hydrochloric acid.

26. Blockage of the bile duct would MOST LIKELY
   A. decrease bile production.
   B. affect the digestion of fat.
   C. raise the pH of the duodenum.
   D. decrease the quantity of feces.
27. Identify the structure labelled X in the above diagram.
   A. Cecum.
   B. Duodenum.
   C. Cardiac Sphincter.
   D. Pyloric sphincter.

28. Which ion is necessary for blood clotting?
   A. Iron.
   B. Sodium.
   C. Calcium.
   D. Chloride.

29. Which of the following are absorbed by lacteals?
   A. Starches.
   B. Fatty acids.
   C. Amino Acids.
   D. Monosaccharides.
30. MOST of the carbon dioxide transported to the lungs by the circulatory system is in the form of

A. carbonic acid.
B. bicarbonate ions.
C. calcium carbonate.
D. carbonic anhydrase.

31. If a person with blood group A receives type AB blood in a transfusion, which of the following occurs?

A. Agglutination.
B. Clotting.
C. Erythroblastosis.
D. No reaction.

32. Which blood vessels have the greatest surface area to volume ratio?

A. Veins.
B. Venules.
C. Arterioles.
D. Capillaries.

33. In a blood pressure reading of 120/80, 80 represents the pressure

A. in the capillaries.
B. in the veins.
C. when the ventricles have relaxed.
D. when the ventricles have contracted.
Use the following diagram to answer question 34.

34. Which of the following is true of blood velocity?

A. Velocity is fastest at Z, slowest at X.
B. Velocity is fastest at Y, slowest at Z.
C. Velocity is fastest at X, slowest at Y.
D. Velocity is fastest at X, slowest at Z.

35. Which of the following is associated with increased heartbeat and breathing rate?

A. Aldosterone.
B. Adrenalin.
C. Acetylcholine.
D. Cholinesterase.

36. What type of neuron transmits an impulse from a receptor to the CNS (central nervous system)?

A. Motor.
B. Efferent.
C. Sensory.
D. Interneuron.

37. The correct pathway of an impulse passing through a reflex arc is

A. receptor, sensory neuron, motor neuron, effector.
B. effector, sensory neuron, motor neuron, receptor.
C. sensory neuron, motor neuron, effector, receptor.
D. motor neuron, receptor, sensory neuron, effector.
38. Which of the following describes an effect of alcohol on an individual?
   A. Narcotic.
   B. Stimulant.
   C. Depressant.
   D. Hallucinogen.

39. The control of heartbeat, breathing, and reflex activities lies with the
   A. cerebrum.
   B. thalamus.
   C. cerebellum.
   D. medulla oblongata.

40. The sodium/potassium pump in neurons is involved with
   A. excretion of salts.
   B. resting potential.
   C. synaptic transmission.
   D. threshold level.

Use the following diagram to answer question 41.

41. During which stage in the above graph is energy used?
   A. W
   B. X
   C. Y
   D. Z
Use the following diagram to answer question 42.

42. The structure labelled Z in the above diagram is
   A. an axon.
   B. a dendrite.
   C. a nucleus.
   D. a node of Ranvier.

43. Receptors in the breathing centre of the medulla oblongata are stimulated by
   A. low O₂ levels.
   B. high O₂ levels.
   C. low CO₂ levels.
   D. high CO₂ levels.

44. Urea is produced in the
   A. liver.
   B. kidney.
   C. pancreas.
   D. large intestine.

45. During expiration (exhalation) the diaphragm
   A. relaxes and the rib cage lifts.
   B. relaxes and the rib cage drops.
   C. contracts and the rib cage lifts.
   D. contracts and the rib cage drops.

46. Which would be an indication of kidney malfunction?
   A. Salts in the loop of Henle.
   B. Glucose in the glomerulus.
   C. Hemoglobin in the Bowman's capsule.
   D. Uric acid in the collecting ducts.
47. In the kidney, nutrients and salt ions are selectively reabsorbed into the
   A. renal artery.
   B. collecting duct.
   C. afferent arteriole.
   D. peritubular capillaries.

48. Fluids appear in the Bowman's capsule as a result of
   A. reabsorption.
   B. blood pressure.
   C. osmotic pressure.
   D. tubular excretion.

49. If a person has had very little to drink on a hot day the pituitary gland would respond by secreting more
   A. adrenalin.
   B. thyroxin.
   C. PTH (parathormone).
   D. ADH (antidiuretic hormone).

50. A hormone is injected into muscle tissue. The muscle cells are observed to increase their rate of cellular respiration. This hormone was MOST LIKELY obtained from the
   A. testes.
   B. thyroid.
   C. pancreas.
   D. posterior pituitary.

51. Hormones travel to their target cells by way of
   A. villi.
   B. ducts.
   C. lacteals.
   D. capillaries.
52. Which of the following BEST displays a graphic representation of a homeostatic mechanism?

A

B

C

D

END OF EXAMINATION
FORM II

NAME: ___________________________ NUMBER: ____________
SCHOOL: __________________________

BIOLOGY 12

GENERAL INSTRUCTIONS

1. Write your name and school in the space provided above.

2. DO NOT WRITE YOUR NAME ON THE ANSWER SHEET.

3. Ensure that the Form Number (I or II) on this booklet and the answer sheet are the same.

4. Ensure that the I.D. Number (1 to 600) on answer sheet is the same as on the exam booklet.

5. When writing the second form (I or II) ensure that the I.D. Number is the SAME I.D. as on the first form. (I or II)

6. When instructed to open this booklet, check the numbering of the pages to ensure that they are numbered in sequence from page 1 to the last page which is identified by END OF EXAMINATION.

7. At the end of the examination, place your answer sheet inside the front cover of this booklet and return the booklet and your answer sheet to the supervisor.

Value: 52 marks (one mark per question) Time: 40 minutes

INSTRUCTIONS:

For each question, select the BEST answer and record your choice on the answer sheet provided. Using an HB pencil, completely fill in the circle that has the letter corresponding to your answer.
1. Which is the correct relationship?

A. DNA at X produces mRNA (messenger RNA) which is used at W.
B. DNA at Y produces mRNA (messenger RNA) which is used at X.
C. DNA at W is used to make protein at X.
D. DNA at X is used to make protein at Z.

2. A carbon compound has 2 hydrogen atoms for every oxygen atom. This compound is a

A. fat.
B. protein.
C. nucleotide.
D. carbohydrate.
Use the following chart to answer question 3.

<table>
<thead>
<tr>
<th>BASE COMPOSITION OF A PIG'S THYMUS CELLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>27%</td>
</tr>
</tbody>
</table>

3. The above chart shows and experimental analysis of the DNA bases contained in the cells of a pig's thymus. The correct identification of the bases X, Y and Z respectively is:
   A. adenine, cytosine, thymine.
   B. thymine, adenine, cytosine.
   C. cytosine, adenine, thymine.
   D. cytosine, thymine, adenine.

4. The anticodon on tRNA is:
   A. identical to the codon on mRNA.
   B. identical to the triplet code on rRNA.
   C. complementary to the codon on mRNA.
   D. complementary to the triplet code on DNA.

5. DNA molecules isolated from onion and human cells differ in their:
   A. type of sugars.
   B. sequence of bases.
   C. number of strands.
   D. order of phosphates.
6. Identify the structure labelled Z in the above diagram.

A. Lysosome.
B. Ribosomes.
C. Golgi apparatus.
D. Endoplasmic reticulum.

7. According to the fluid mosaic model of the cell membrane,

A. proteins are scattered throughout a double layer of phospholipid.
B. proteins are sandwiched between two layers of cellulose.
C. phospholipids are sandwiched between two layers of protein.
D. phospholipids are scattered throughout a double layer of cellulose.
8. In the above experimental situation, osmosis is the term used to describe the movement of

A. salt from X to Y.
B. salt from Y to X.
C. water from Y to X.
D. water from X to Y.

9. Which cellular function would be interfered with if lysosomes were destroyed?

A. RNA synthesis.
B. Cell secretion.
C. Cell digestion.
D. Packaging of molecules.

10. What kind of chemical reaction does NAD → NADH₂ represent?

A. Reduction.
B. Synthesis.
C. Oxidation.
D. Phosphorylation.
11. With reference to the above diagram, when the amount of "Z" increases in a cell, enzyme "1" will convert

A. less X to Y.
B. less Y to X.
C. more X to Y.
D. more Y to X.

12. The cell process which uses energy to bring substances into the cell is

A. osmosis.
B. diffusion.
C. active transport.
D. facilitated transport.

13. With all other factors constant, which of the following will increase the quantity of the end product of an enzyme-catalysed reaction?

A. Increase the temperature to 100 C.
B. Increase the amount of the substrate.
C. Maintain the pH.
D. Maintain the reaction at body temperature.

14. The Calvin cycle reactions in photosynthesis produce

A. ATP.
B. water.
C. oxygen.
D. carbohydrates.
15. According to the above graph, which of the following conditions must be kept constant?

A. Temperature and relative humidity.
B. Water supply and oxygen concentration.
C. Light intensity and carbon dioxide concentration.
D. Oxygen and carbon dioxide concentration.

16. Substances which may contribute atoms to an enzyme-catalyzed reaction are called

A. inhibitors.
B. coenzymes.
C. apoenzymes.
D. heavy metals.

17. The Calvin cycle in photosynthesis uses energy from

A. NADPH\textsubscript{2} produced only in the thylakoids.
B. NADPH\textsubscript{2} and ATP produced in the thylakoids.
C. NADPH\textsubscript{2} and ATP produced in the stroma.
D. NADPH\textsubscript{2} produced only in the stroma.
18. A lowered concentration of environmental carbon dioxide would affect a plant by
   A. raising its rate of respiration.
   B. lowering its rate of respiration.
   C. raising its rate of photosynthesis.
   D. lowering its rate of photosynthesis.

19. During the light-capturing of photosynthesis, light energy is
   A. used to make glucose.
   B. used to make carbon dioxide.
   C. changed into three carbon sugars.
   D. transformed into chemical energy.

20. Which one of the following is an example of a tissue?
    A. Eye.
    B. Skin.
    C. Blood.
    D. Pancreas.

21. Blockage of the substances leaving the pancreas would
    A. decrease bile production.
    B. raise the pH of the duodenum.
    C. decrease the quantity of feces.
    D. affect the digestion of protein.

22. The conversion of amino acids to glucose occurs mainly in the
    A. liver.
    B. thymus.
    C. spleen.
    D. pancreas.

23. Which of the following would produce the greatest amount of energy per gram?
    A. Fat.
    B. Sugar.
    C. Starch.
    D. Protein.

24. The correct sequence of structures that food contacts as it moves along the digestive system is
    A. mouth, stomach, large intestine, small intestine, anus.
    B. pharynx, small intestine, stomach, large intestine, anus.
    C. mouth, esophagus, stomach, small intestine, large intestine.
    D. esophagus, pharynx, stomach, large intestine, small intestine.
25. The substance secreted by the walls of the stomach that prevents the stomach from digesting itself is
   A. water.
   B. mucus.
   C. pepsin.
   D. gastrin.

26. Which digestive enzyme is incorrectly matched to its substrate?
   A. Lipase - fat.
   B. Pepsin - protein.
   C. Trypsin - nucleic acid.
   D. Salivary amylase - starch.

Use the following information to answer question 27.

27. Which number represents the organ that secretes amylase?
   A. 1.
   B. 2.
   C. 3.
   D. 4.
28. Deoxygenated hemoglobin is found in its greatest concentration in
A. pulmonary veins.
B. systemic arterioles.
C. pulmonary arteries.
D. coronary arteries.

29. Failure of the lymphatic system to collect tissue fluids results in
A. infection.
B. tissue swelling.
C. excessive urination.
D. growth of adipose tissue.

30. Which of the following are necessary for the production of thrombin
in an area of injured tissue?
A. Fibrin and calcium ions.
B. Fibrin and thromboplastin.
C. Calcium ions and prothrombin.
D. Fibrinogen and thromboplastin.

31. A blood pressure reading of 165/100 is characteristic of
A. hypotension.
B. hypertension.
C. a normal resting adult.
D. insufficient dietary salt.

32. Blood plasma with fibrinogen removed is known as
A. serum.
B. lymph.
C. formed elements.
D. tissue fluid.
33. Which blood vessel is labelled X in the above diagram?
   A. Anterior vena cava.
   B. Pulmonary trunk.
   C. Pulmonary vein.
   D. Aorta.

34. In the fetus, blood passes from the right atrium to the left atrium through the
   A. semi-lunar valves.
   B. ductus venosus (venous duct).
   C. foramen ovale (oval opening).
   D. ductus arteriosus (arterial duct).

35. Which part of the nervous system regulates body temperature?
   A. Thalamus.
   B. Cerebellum.
   C. Hypothalamus.
   D. Cerebral cortex.

36. A sensory neuron carries information
   A. to muscles.
   B. from the central nervous system.
   C. to the central nervous system.
   D. between interneurons.
37. When frightened, a person's pupil size increases. The part of the nervous system that controls this response is the
   A. sensory nervous system.
   B. somatic nervous system.
   C. sympathetic nervous system.
   D. parasympathetic nervous system.

38. Sodium ions are moved out of a neuron by
   A. passive transport.
   B. diffusion.
   C. osmosis.
   D. active transport.

39. The type of sensation experienced as a result of a nerve impulse depends on the
   A. area of the brain stimulated.
   B. part of the spinal cord that is stimulated.
   C. strength of the impulse.
   D. type of effector involved.

40. Damage to the occipital lobes of the brain may result in impaired
   A. smell.
   B. speech.
   C. vision.
   D. hearing.
41. Which number represents an effector?
   A. 1  
   B. 2  
   C. 3  
   D. 4

42. Which of the following causes human lungs to inflate?
   A. Contraction of abdominal muscles.  
   B. Contraction of the diaphragm.  
   C. Relaxation of the diaphragm.  
   D. Relaxation of intercostal muscles.

43. Below the pharynx, which is the correct descending anatomical order?
   A. Larynx, epiglottis, trachea, alveoli.  
   B. Epiglottis, trachea, larynx, alveoli.  
   C. Epiglottis, larynx, trachea, alveoli.  
   D. Larynx, trachea, epiglottis, alveoli.

44. If airplane passengers were to experience a lower level of atmospheric oxygen, what response would occur immediately?
   A. The heart rate would decrease.  
   B. Calcitonin levels would increase.  
   C. The breathing rate would increase.  
   D. More hemoglobin would be produced.
45. After prolonged sweating and little water intake, your body would respond by
   A. secreting more ADH.
   B. reducing aldosterone levels.
   C. increasing urine production.
   D. decreasing collecting duct permeability.

46. Which of the following is the BEST indicator of abnormal kidney function?
   A. Glucose in the urine.
   B. Protein in the urine.
   C. Hormones in the urine.
   D. Electrolytes in the urine.

47. Hemoglobin pigments are excreted by the
   A. liver.
   B. marrow.
   C. spleen.
   D. pancreas.

Use the following diagram to answer question 48.

48. Pressure filtration occurs at the region labelled
   A. 1
   B. 2
   C. 3
   D. 4
49. Damage to the posterior pituitary gland would have an effect on
   A. adrenal glands.
   B. kidney function.
   C. the pancreas.
   D. the thyroid gland.

50. Which of the following glands has both an endocrine and an exocrine function?
   A. Adrenals.
   B. Thyroid.
   C. Pituitary.
   D. Pancreas.

51. Deamination of amino acids produces ammonia which is then converted to
   A. urea.
   B. protein.
   C. creatinine.
   D. bile.

52. Which hormone is a secretion of the adrenal medulla?
   A. Cortisol.
   B. Adrenalin.
   C. Aldosterone.
   D. Testosterone.

END OF EXAMINATION
Appendix D

Request and Permission Forms
CONSENT FORM

I, ________________ Superintendent of School District
______________ consent to the involvement
of Biology 12 students in the examination equivalency study as
outlined in the letter from Peter MacMillan.

Peter MacMillan has permission to contact any or all Biology
12 in the district requesting their participation. It is the
teachers' professional decision as to whether or not they wish
their classes to participate.

__________________________
Appendix E

Teacher Administration Instructions
ADMINISTRATION INSTRUCTIONS

TO THE TEACHER

Advance Information

Students may be told ahead of time that they will be writing two forms of an examination that is equivalent to the multiple choice section of the Biology 12 government examination.

Students should be told one class in advance of when an exam writing will occur.

Students may be given any information about the study that you wish to give BUT only after the second form has been written.

If you or the students wish any further information or have any comments about the study, please send the requests or comments to me when the answer sheets are being returned.

Prior to the Examination

Please check that you have equal numbers of FORM I and FORM II of the examination. Feel free to xerox extra exam copies if necessary. Extra answer sheets have been included with the package.
Your school has been selected by lot to write either FORM I or FORM II first. Please ensure that all students in all classes in your school write the two forms in the sequence selected for you.

Please attempt to coordinate the classes in which the forms are written so that discussion between classes is minimized.

The other form should be written as soon as feasible after the first form; back to back or within two to three class periods would be ideal.

At the Time of Administration

Give the students answer sheets and tell them TO PUT NAMES ON THEM BUT NOT TO BUBBLE IN THEIR NAMES.

Review directions for marking bubble sheets found on side 2 of the answer sheet.

Have students bubble in the sex, grade, birth date columns. DO NOT bubble in any I.D. number even if you are intending to mark the sheets using a scanner in your school.

IF YOUR SCHOOL IS SELECTED TO WRITE FORM I FIRST, BUBBLE IN THE '0' BUBBLE OF SPECIAL CODE K. IF YOU ARE WRITING FORM II FIRST, BUBBLE IN THE '0' BUBBLE OF SPECIAL CODE L.
ADMINISTRATION INSTRUCTIONS

WHEN WRITING FORM I BUBBLE IN ANSWERS 1 TO 52. WHEN WRITING FORM II BUBBLE IN ANSWERS 101 TO 152.

The students should read the Student Instruction Sheet replacing general instructions 2. to 5. when appropriate.

All other rules for examination writing shall be the same as for the Biology 12 government examination e.g. calculator usage, scrap paper, teacher non-assistance. The exception being that the students are supervised by their classroom teacher.

After Administration of First Form

Collect all test booklets with answer sheets placed in the booklets.

Check for name but remove any bubbled in names on answer sheets.

Separate and store examination booklets and answer sheets in a secure location.

DO NOT RETURN, OR MARK, OR DISCUSS THE EXAM AT THIS TIME!

Administration of the Second Form

Return the student answer sheets to the original owners.

Follow the same administration procedure as for the first form.
Appendix F

Equipercntile Equating Tables
### SCOREA

<table>
<thead>
<tr>
<th>VALUE</th>
<th>LABEL</th>
<th>VALUE</th>
<th>FREQUENCY</th>
<th>PERCENT</th>
<th>VALID PERCENT</th>
<th>CUM PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td></td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1.4</td>
<td>3.1</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>1.4</td>
<td>3.1</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>4</td>
<td>4</td>
<td>1.4</td>
<td>1.4</td>
<td>3.1</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>7</td>
<td>7</td>
<td>2.4</td>
<td>2.4</td>
<td>3.1</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>5</td>
<td>1.7</td>
<td>1.7</td>
<td>4.2</td>
<td>4.2</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>5</td>
<td>1.7</td>
<td>1.7</td>
<td>8.4</td>
<td>8.4</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>9</td>
<td>3.1</td>
<td>3.1</td>
<td>11.5</td>
<td>11.5</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>15</td>
<td>5.2</td>
<td>5.2</td>
<td>16.8</td>
<td>16.8</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>9</td>
<td>3.1</td>
<td>3.1</td>
<td>19.9</td>
<td>19.9</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>7</td>
<td>2.4</td>
<td>2.4</td>
<td>22.4</td>
<td>22.4</td>
</tr>
<tr>
<td>21</td>
<td></td>
<td>16</td>
<td>5.6</td>
<td>5.6</td>
<td>28.0</td>
<td>28.0</td>
</tr>
<tr>
<td>22</td>
<td></td>
<td>11</td>
<td>3.8</td>
<td>3.8</td>
<td>31.8</td>
<td>31.8</td>
</tr>
<tr>
<td>23</td>
<td></td>
<td>8</td>
<td>2.8</td>
<td>2.8</td>
<td>34.6</td>
<td>34.6</td>
</tr>
<tr>
<td>24</td>
<td></td>
<td>9</td>
<td>3.1</td>
<td>3.1</td>
<td>37.8</td>
<td>37.8</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>11</td>
<td>3.8</td>
<td>3.8</td>
<td>41.6</td>
<td>41.6</td>
</tr>
<tr>
<td>26</td>
<td></td>
<td>9</td>
<td>3.1</td>
<td>3.1</td>
<td>44.8</td>
<td>44.8</td>
</tr>
<tr>
<td>27</td>
<td></td>
<td>9</td>
<td>3.1</td>
<td>3.1</td>
<td>47.9</td>
<td>47.9</td>
</tr>
<tr>
<td>28</td>
<td></td>
<td>9</td>
<td>3.1</td>
<td>3.1</td>
<td>51.0</td>
<td>51.0</td>
</tr>
<tr>
<td>29</td>
<td></td>
<td>14</td>
<td>4.9</td>
<td>4.9</td>
<td>55.9</td>
<td>55.9</td>
</tr>
<tr>
<td>30</td>
<td></td>
<td>6</td>
<td>2.1</td>
<td>2.1</td>
<td>58.0</td>
<td>58.0</td>
</tr>
<tr>
<td>31</td>
<td></td>
<td>12</td>
<td>4.2</td>
<td>4.2</td>
<td>62.2</td>
<td>62.2</td>
</tr>
<tr>
<td>32</td>
<td></td>
<td>10</td>
<td>3.5</td>
<td>3.5</td>
<td>65.7</td>
<td>65.7</td>
</tr>
<tr>
<td>33</td>
<td></td>
<td>8</td>
<td>2.8</td>
<td>2.8</td>
<td>68.5</td>
<td>68.5</td>
</tr>
<tr>
<td>34</td>
<td></td>
<td>12</td>
<td>4.2</td>
<td>4.2</td>
<td>72.7</td>
<td>72.7</td>
</tr>
<tr>
<td>35</td>
<td></td>
<td>5</td>
<td>1.7</td>
<td>1.7</td>
<td>74.5</td>
<td>74.5</td>
</tr>
<tr>
<td>36</td>
<td></td>
<td>3</td>
<td>1.0</td>
<td>1.0</td>
<td>75.5</td>
<td>75.5</td>
</tr>
<tr>
<td>37</td>
<td></td>
<td>7</td>
<td>2.4</td>
<td>2.4</td>
<td>78.0</td>
<td>78.0</td>
</tr>
<tr>
<td>38</td>
<td></td>
<td>8</td>
<td>2.8</td>
<td>2.8</td>
<td>80.8</td>
<td>80.8</td>
</tr>
<tr>
<td>39</td>
<td></td>
<td>9</td>
<td>3.1</td>
<td>3.1</td>
<td>83.9</td>
<td>83.9</td>
</tr>
<tr>
<td>40</td>
<td></td>
<td>6</td>
<td>2.1</td>
<td>2.1</td>
<td>86.0</td>
<td>86.0</td>
</tr>
<tr>
<td>41</td>
<td></td>
<td>8</td>
<td>2.8</td>
<td>2.8</td>
<td>88.8</td>
<td>88.8</td>
</tr>
<tr>
<td>42</td>
<td></td>
<td>7</td>
<td>2.4</td>
<td>2.4</td>
<td>91.3</td>
<td>91.3</td>
</tr>
<tr>
<td>43</td>
<td></td>
<td>2</td>
<td>0.7</td>
<td>0.7</td>
<td>92.0</td>
<td>92.0</td>
</tr>
<tr>
<td>44</td>
<td></td>
<td>6</td>
<td>2.1</td>
<td>2.1</td>
<td>94.1</td>
<td>94.1</td>
</tr>
<tr>
<td>45</td>
<td></td>
<td>5</td>
<td>1.7</td>
<td>1.7</td>
<td>95.8</td>
<td>95.8</td>
</tr>
<tr>
<td>46</td>
<td></td>
<td>4</td>
<td>1.4</td>
<td>1.4</td>
<td>97.2</td>
<td>97.2</td>
</tr>
<tr>
<td>47</td>
<td></td>
<td>2</td>
<td>0.7</td>
<td>0.7</td>
<td>97.9</td>
<td>97.9</td>
</tr>
<tr>
<td>48</td>
<td></td>
<td>5</td>
<td>1.7</td>
<td>1.7</td>
<td>99.7</td>
<td>99.7</td>
</tr>
<tr>
<td>51</td>
<td></td>
<td>1</td>
<td>0.3</td>
<td>0.3</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

**TOTAL** 286 100.0 100.0

<table>
<thead>
<tr>
<th>PERCENTILE</th>
<th>VALUE</th>
<th>PERCENTILE VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>11.740</td>
<td>100.00</td>
</tr>
</tbody>
</table>

**VALID CASES** 286  **MISSING CASES** 0
## Scoreb

<table>
<thead>
<tr>
<th>Value Label</th>
<th>Value</th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cum Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>3</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>3</td>
<td>3.0</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>1.0</td>
<td>1.0</td>
<td>2.4</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>1.4</td>
<td>1.4</td>
<td>3.8</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>3</td>
<td>1.0</td>
<td>1.0</td>
<td>4.9</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>7</td>
<td>2.4</td>
<td>2.4</td>
<td>7.3</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>15</td>
<td>5.2</td>
<td>5.2</td>
<td>12.6</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>11</td>
<td>3.8</td>
<td>3.8</td>
<td>16.4</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>10</td>
<td>3.5</td>
<td>3.5</td>
<td>19.9</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>9</td>
<td>3.1</td>
<td>3.1</td>
<td>23.1</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>15</td>
<td>5.2</td>
<td>5.2</td>
<td>28.3</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>19</td>
<td>6.6</td>
<td>6.6</td>
<td>35.0</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>6</td>
<td>2.1</td>
<td>2.1</td>
<td>37.1</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>9</td>
<td>3.1</td>
<td>3.1</td>
<td>40.2</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>12</td>
<td>4.2</td>
<td>4.2</td>
<td>44.4</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>7</td>
<td>2.4</td>
<td>2.4</td>
<td>46.9</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>9</td>
<td>3.1</td>
<td>3.1</td>
<td>50.0</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>11</td>
<td>3.8</td>
<td>3.8</td>
<td>53.8</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>10</td>
<td>3.5</td>
<td>3.5</td>
<td>57.3</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>14</td>
<td>4.9</td>
<td>4.9</td>
<td>62.2</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>7</td>
<td>2.4</td>
<td>2.4</td>
<td>64.7</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>5</td>
<td>1.7</td>
<td>1.7</td>
<td>66.4</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>15</td>
<td>5.2</td>
<td>5.2</td>
<td>71.7</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>11</td>
<td>3.8</td>
<td>3.8</td>
<td>75.5</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>3</td>
<td>1.0</td>
<td>1.0</td>
<td>76.6</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>10</td>
<td>3.5</td>
<td>3.5</td>
<td>80.1</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>6</td>
<td>2.1</td>
<td>2.1</td>
<td>82.2</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>9</td>
<td>3.1</td>
<td>3.1</td>
<td>85.3</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>8</td>
<td>2.8</td>
<td>2.8</td>
<td>88.1</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>7</td>
<td>2.4</td>
<td>2.4</td>
<td>90.6</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>5</td>
<td>1.7</td>
<td>1.7</td>
<td>92.3</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>2</td>
<td>0.7</td>
<td>0.7</td>
<td>93.0</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>3</td>
<td>1.0</td>
<td>1.0</td>
<td>94.1</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>3</td>
<td>1.0</td>
<td>1.0</td>
<td>95.1</td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>5</td>
<td>1.7</td>
<td>1.7</td>
<td>96.9</td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>4</td>
<td>1.4</td>
<td>1.4</td>
<td>98.3</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>2</td>
<td>0.7</td>
<td>0.7</td>
<td>99.0</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>1</td>
<td>0.3</td>
<td>0.3</td>
<td>99.3</td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>2</td>
<td>0.7</td>
<td>0.7</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

**Total** 286 100.0 100.0

<table>
<thead>
<tr>
<th>Percentile</th>
<th>Value</th>
<th>Percentile</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>12.000</td>
<td>100.00</td>
<td></td>
</tr>
</tbody>
</table>

**Valid Cases** 286  **Missing Cases** 0
<table>
<thead>
<tr>
<th>Score A</th>
<th>Score B*</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-8</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>11</td>
<td>-2</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
<td>-2</td>
</tr>
<tr>
<td>11</td>
<td>13</td>
<td>-2</td>
</tr>
<tr>
<td>12</td>
<td>14</td>
<td>-2</td>
</tr>
<tr>
<td>13</td>
<td>15</td>
<td>-2</td>
</tr>
<tr>
<td>14</td>
<td>16</td>
<td>-2</td>
</tr>
<tr>
<td>15</td>
<td>17</td>
<td>-2</td>
</tr>
<tr>
<td>16</td>
<td>18</td>
<td>-2</td>
</tr>
<tr>
<td>17</td>
<td>19</td>
<td>-2</td>
</tr>
<tr>
<td>18</td>
<td>20</td>
<td>-2</td>
</tr>
<tr>
<td>19</td>
<td>21</td>
<td>-2</td>
</tr>
<tr>
<td>20</td>
<td>22</td>
<td>-2</td>
</tr>
<tr>
<td>21</td>
<td>23</td>
<td>-2</td>
</tr>
<tr>
<td>22</td>
<td>24</td>
<td>-2</td>
</tr>
<tr>
<td>23</td>
<td>24</td>
<td>-1</td>
</tr>
<tr>
<td>24</td>
<td>25</td>
<td>-1</td>
</tr>
<tr>
<td>25</td>
<td>26</td>
<td>-1</td>
</tr>
<tr>
<td>26</td>
<td>27</td>
<td>-1</td>
</tr>
<tr>
<td>27</td>
<td>28</td>
<td>-1</td>
</tr>
<tr>
<td>28</td>
<td>29</td>
<td>-1</td>
</tr>
<tr>
<td>29</td>
<td>30</td>
<td>-1</td>
</tr>
<tr>
<td>30</td>
<td>31</td>
<td>-1</td>
</tr>
<tr>
<td>31</td>
<td>32</td>
<td>-1</td>
</tr>
<tr>
<td>32</td>
<td>33</td>
<td>-1</td>
</tr>
<tr>
<td>33</td>
<td>34</td>
<td>-1</td>
</tr>
<tr>
<td>34</td>
<td>35</td>
<td>-1</td>
</tr>
<tr>
<td>35</td>
<td>36</td>
<td>-1</td>
</tr>
<tr>
<td>36</td>
<td>37</td>
<td>-1</td>
</tr>
<tr>
<td>37</td>
<td>37</td>
<td>-1</td>
</tr>
<tr>
<td>38</td>
<td>38</td>
<td>0</td>
</tr>
<tr>
<td>39</td>
<td>39</td>
<td>0</td>
</tr>
<tr>
<td>40</td>
<td>40</td>
<td>0</td>
</tr>
<tr>
<td>41</td>
<td>41</td>
<td>0</td>
</tr>
<tr>
<td>42</td>
<td>42</td>
<td>0</td>
</tr>
<tr>
<td>43</td>
<td>43</td>
<td>0</td>
</tr>
<tr>
<td>44</td>
<td>45</td>
<td>-1</td>
</tr>
<tr>
<td>45</td>
<td>46</td>
<td>-1</td>
</tr>
<tr>
<td>46</td>
<td>47</td>
<td>-1</td>
</tr>
<tr>
<td>47</td>
<td>48</td>
<td>-1</td>
</tr>
<tr>
<td>48</td>
<td>49</td>
<td>-1</td>
</tr>
<tr>
<td>49</td>
<td>50</td>
<td>-1</td>
</tr>
<tr>
<td>50</td>
<td>51</td>
<td>-1</td>
</tr>
<tr>
<td>51</td>
<td>52</td>
<td>-1</td>
</tr>
<tr>
<td>CP</td>
<td>Score</td>
<td>CP</td>
</tr>
<tr>
<td>----</td>
<td>--------</td>
<td>----</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>1</td>
<td>11.0</td>
<td>25</td>
</tr>
<tr>
<td>2</td>
<td>12.5</td>
<td>25</td>
</tr>
<tr>
<td>3</td>
<td>13.2</td>
<td>25</td>
</tr>
<tr>
<td>4</td>
<td>13.8</td>
<td>25</td>
</tr>
<tr>
<td>5</td>
<td>14.4</td>
<td>26</td>
</tr>
<tr>
<td>6</td>
<td>15.0</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>15.3</td>
<td>26</td>
</tr>
<tr>
<td>8</td>
<td>15.8</td>
<td>26</td>
</tr>
<tr>
<td>9</td>
<td>16.2</td>
<td>26</td>
</tr>
<tr>
<td>10</td>
<td>16.6</td>
<td>27</td>
</tr>
<tr>
<td>11</td>
<td>16.9</td>
<td>27</td>
</tr>
<tr>
<td>12</td>
<td>17.2</td>
<td>27</td>
</tr>
<tr>
<td>13</td>
<td>17.4</td>
<td>28</td>
</tr>
<tr>
<td>14</td>
<td>17.7</td>
<td>28</td>
</tr>
<tr>
<td>15</td>
<td>17.9</td>
<td>28</td>
</tr>
<tr>
<td>16</td>
<td>18.1</td>
<td>29</td>
</tr>
<tr>
<td>17</td>
<td>18.3</td>
<td>29</td>
</tr>
<tr>
<td>18</td>
<td>18.5</td>
<td>29</td>
</tr>
<tr>
<td>19</td>
<td>18.9</td>
<td>29</td>
</tr>
<tr>
<td>20</td>
<td>19.1</td>
<td>29</td>
</tr>
<tr>
<td>21</td>
<td>19.5</td>
<td>30</td>
</tr>
<tr>
<td>22</td>
<td>19.7</td>
<td>30</td>
</tr>
<tr>
<td>23</td>
<td>19.9</td>
<td>30</td>
</tr>
<tr>
<td>24</td>
<td>20.2</td>
<td>31</td>
</tr>
<tr>
<td>25</td>
<td>20.5</td>
<td>31</td>
</tr>
<tr>
<td>26</td>
<td>20.7</td>
<td>31</td>
</tr>
<tr>
<td>27</td>
<td>20.9</td>
<td>31</td>
</tr>
<tr>
<td>28</td>
<td>21.1</td>
<td>32</td>
</tr>
<tr>
<td>29</td>
<td>21.4</td>
<td>32</td>
</tr>
<tr>
<td>30</td>
<td>21.7</td>
<td>32</td>
</tr>
<tr>
<td>31</td>
<td>21.9</td>
<td>33</td>
</tr>
<tr>
<td>32</td>
<td>22.2</td>
<td>33</td>
</tr>
<tr>
<td>33</td>
<td>22.3</td>
<td>34</td>
</tr>
<tr>
<td>34</td>
<td>22.8</td>
<td>24</td>
</tr>
</tbody>
</table>