

TEST-WISENESS: ITS EFFECT ON THE SUPPLY ITEMS OF THE  
BRITISH COLUMBIA PROVINCIAL EXAMINATIONS FOR GRADE  
TWELVE STUDENTS

By

Michelle Mae Vanchu

B. Ed. (Special Education) University of British Columbia, 1983

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

in

THE FACULTY OF GRADUATE STUDIES  
EDUCATIONAL PSYCHOLOGY AND SPECIAL EDUCATION

We accept this thesis as conforming  
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August 1990

© Michelle Mae Vanchu, 1990

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 Special Education  
Vancouver, Canada

Date August 16/90

## **Abstract**

Test-wiseness, possessed in different amounts by different individuals, is the ability to use test format, test characteristics, and/or the testing situation to receive a high score. As such, test-wiseness is an unwanted source of variance which can inflate test scores, thus invalidating test results. Problems of inappropriate interpretation may arise when test scores are affected by test-wiseness.

The present study addressed the relationship between test-wiseness and English ability, as measured by the British Columbia Provincial English 12 Examination for June of 1989. The English 12 examination contained both selection and supply items. This provided an opportunity to examine both types of items and their relationship to test-wiseness. Previous research had focussed on selection items.

To provide a framework for understanding and presenting the results the present research was divided into two substudies.

Substudy I addressed questions concerning the nature and strength of the relationship of test-wiseness to the selection, short-answer, and extended-answer items of the English 12 examination. The selection items of the English 12 examination provided reference for interpreting the results for the short-answer and extended-answer items. Test scores were adjusted for the presence of verbal ability and it was found, as previous research indicated, that test-wiseness and verbal ability were moderately correlated.

To further clarify the concept of test-wiseness, differences between test-wise and test-naïve students were examined in terms of means and variability on selection, short-answer, and extended-answer items of the English 12 examination.

The results of the study are based on test data for 735 students collected from April

to June of 1989. Each student completed the Test of Test-Wiseness (TTW), Language Proficiency Index (LPI), and a form containing ethnographic information. Test scores for the English 12 examination were provided by the Ministry.

Based on the analyses of test data for 735 grade twelve students, test-wiseness accounted for less than four percent of the variance on the English 12 examination for selection, short-answer, and extended-answer items on Substudy I. These results were found to be statistically significant. Results for Substudy II indicated that there were differences between test-wise and test-naive students in terms of means for the selection and short-answer items. Results for the extended-answer (essay) item were non-significant. There were no differences in variability between the test-wise and test-naive samples for any of the item types.

The results of the present study will be of interest to those involved in constructing the English 12 examination, as well as grade 12 teachers and students. The test score on the English 12 examination accounts for 40% of a student's English 12 final grade, with a student's graduation or failure based upon these results. As such, English 12 examination scores should be as accurate and valid as possible.

## Table of Contents

<b>Abstract</b>	<b>ii</b>
<b>List of Tables</b>	<b>viii</b>
<b>Acknowledgment</b>	<b>ix</b>
<b>1 The Problem In Its Setting</b>	<b>1</b>
1.1 Test-Wiseness . . . . .	2
1.2 The British Columbia Provincial Examinations . . . . .	5
1.3 Statement of the Problem . . . . .	6
1.4 Definitions Used In This Study . . . . .	8
1.5 Selection of the Dependent Variable . . . . .	9
1.6 Hypotheses of the Study . . . . .	10
1.7 Organization of the Thesis . . . . .	13
<b>2 Review of the Literature</b>	<b>14</b>
2.1 Test-Wiseness . . . . .	14
2.1.1 Definition(s) of Test-Wiseness . . . . .	14
2.1.2 Test-Wiseness and Bias . . . . .	18
2.1.3 The Nature of Test-Wiseness . . . . .	19
2.2 Test-Wiseness Research . . . . .	21
2.2.1 Correlates of Test-Wiseness . . . . .	21
2.2.2 Developmental Nature of Test-Wiseness . . . . .	23

2.2.3	Variability of Test-Wiseness Ability . . . . .	24
2.2.4	The Effect of Test-wiseness On Supply or Open-ended Test Items . . . . .	25
2.3	British Columbia Provincial Examination Program . . . . .	25
2.4	Grade Twelve Provincial Examinations: English 12 Examination . . . . .	28
2.4.1	Scoring: The English 12 Examination . . . . .	29
<b>3</b>	<b>Methodology</b>	<b>32</b>
3.1	Sample . . . . .	32
3.2	Instrumentation . . . . .	33
3.2.1	English 12 Examination . . . . .	33
3.2.2	Test of Test-Wiseness . . . . .	33
3.2.3	Language Proficiency Index . . . . .	35
3.3	Procedures . . . . .	37
3.3.1	Testing . . . . .	37
3.4	Scoring and Data Entry . . . . .	38
3.5	Statistical Analysis . . . . .	39
3.5.1	Preliminary Analyses . . . . .	39
3.5.2	Description of the Sample . . . . .	39
<b>4</b>	<b>Results</b>	<b>42</b>
4.1	Substudy I . . . . .	42
4.1.1	Description of the Sample . . . . .	42
4.1.2	Test Characteristics . . . . .	45
4.1.3	Results of the Statistical Analyses . . . . .	47
4.2	Substudy II . . . . .	50
4.2.1	Description of the Test-Wise and Test-Naive Samples . . . . .	50
4.2.2	Results of the Statistical Analyses . . . . .	53

<b>5</b>	<b>Conclusions and Recommendations</b>	<b>57</b>
5.1	Summary of the Problem and Procedure . . . . .	57
5.2	Summary and Discussion of the Results . . . . .	58
5.2.1	Substudy I . . . . .	58
5.2.2	Substudy II . . . . .	59
5.3	Limitations . . . . .	60
5.4	Further Discussion . . . . .	60
5.5	Recommendations . . . . .	62
5.5.1	Implications for Practice . . . . .	62
5.5.2	Further Research . . . . .	63
	<b>Bibliography</b>	<b>66</b>
	<b>appendices</b>	<b>71</b>
<b>A</b>	<b>The English Twelve Provincial Examination</b>	<b>71</b>
<b>B</b>	<b>English 12 Composition – Scoring Guide</b>	<b>86</b>
<b>C</b>	<b>Rules for Holistic Scoring</b>	<b>90</b>
<b>D</b>	<b>Test of Test-Wiseness</b>	<b>92</b>
<b>E</b>	<b>Instrument Used For Validation of Test-Wiseness</b>	<b>97</b>
<b>F</b>	<b>Language Proficiency Index</b>	<b>104</b>
<b>G</b>	<b>Protocol for Administration of the Test of Test-Wiseness</b>	<b>106</b>
<b>H</b>	<b>Protocol for Administration of the Language Proficiency Index</b>	<b>108</b>

<b>I Preliminary Analysis: The Unit of Analysis using ONEWAY</b>	<b>109</b>
<b>J Scatterplots for the Standardized Residuals (n=735)</b>	<b>110</b>
<b>K Scatterplots for the Standardized Residuals (n=137)</b>	<b>112</b>



## List of Tables

1.1	Taxonomy of Test-Wiseness Principles A . . . . .	3
1.2	Taxonomy of Test-Wiseness Principles B . . . . .	4
2.3	English 12 Examination, June 1989 . . . . .	29
3.4	Description of the Test of Test-Wiseness . . . . .	34
3.5	Description of the Test-Wiseness Measure Used for Validation . . . . .	36
3.6	Testing Schedule . . . . .	37
4.7	Description of the Full Sample . . . . .	44
4.8	Means, Standard Deviations, and Internal Consistancies . . . . .	46
4.9	Zero-Order Correlation Matrix (n=735) . . . . .	48
4.10	Semi-Partial Correlations . . . . .	48
4.11	Test of Significance of the Semi-Partial Correlation . . . . .	49
4.12	Description of the Test-Wise (n=61) and Test-Naive (n=76) Samples . . . . .	51
4.13	Results of the ANCOVA Analyses for the English 12 Exam . . . . .	54
4.14	Adujsted Means and Standard Deviations . . . . .	55
I.15	Results of ONEWAY Analyses Between Schools for the TTW and LPI . . . . .	109

## Acknowledgment

After spending a large portion of the last three years “busy studying” I would like to thank those people whose support and guidance allowed me to grow educationally, professionally, and personally.

- My family, especially my husband, David, who have given me the support, as well as often needed extra hands, I needed to finish my work.
- Todd Rogers, Dave Bateson, and Bob Conry who have given me advice and encouragement which allowed me to reach within myself and produce work I am proud of.
- Bob Bruce, Stewart Seidel, and Ken Chan who saw me through file merges, combining, and sorting data. Their expertise was invaluable to improving my computer literacy.

## Chapter 1

### The Problem In Its Setting

In 1983, the British Columbia Ministry of Education reintroduced provincial examinations, in thirteen grade 12 subjects, to be written by every grade twelve student who wished to obtain credit for the course. Comprised of multiple choice, completion, and short essay questions, performance on these tests initially counted for one-half of a student's final grade in each examinable course. However, in response to public and professional reactions (see LTAS, p.15) and suggestions contained in the report of the Royal Commission on Education completed in 1988, the weighting for the provincial examinations was changed to 40% of the final grade beginning in the 1989 - 90 school year.

The problem addressed by this study concerned the degree to which test-wiseness influenced the interpretation of the test scores received by students' on the Grade 12 Provincial Examinations. Of particular interest was the impact of test-wiseness upon the performance on the supply items of the English 12 examination.

Possessed in differing amounts by different students, test-wiseness is "widely recognized as a source of additional variance in test scores and as a possible depressor of test [score] validity" (Sarnacki, 1979, p.253). Of concern was the possibility that test-wise students may have an unfair advantage on these examinations, particularly if the items in the examinations are susceptible to test-wiseness. A preliminary study (Rogers & Bateson, 1990a) indicated that some of the test items on the provincial examinations are, indeed, susceptible to test-wiseness.

## 1.1 Test-Wiseness

First formally introduced by Thorndike (1951), test-wiseness refers to the ability of some students to obtain higher test scores based upon certain cues present in the items included in a test. Millman, Bishop, and Ebel (1965) defined test-wiseness as “a subjects’ capacity to utilize characteristics and formats of the test and/or test taking situation to receive a high score” (p.707). Millman et al. further considered this ability to be logically independent of the subjects’ knowledge of the content being tested. In an earlier study of test-wiseness, Gibb (1964), concluded that, like Millman et al., “...test-wiseness may exist separately from subject-matter knowledge; that test-wiseness represents a source of invalid variance; and that there is a significant spread of reliable individual differences in test-wiseness” (p.66).

To explicate their definition of test-wiseness further, Millman et al. (1965) developed the Taxonomy of Test-Wiseness Principles presented in Table 1.1 and Table 1.2. As shown, the elements of this Taxonomy are divided into two basic groups. The first contains elements which are characteristic of the test taker: use of error-avoidance strategies, time-using strategies, guessing strategies, and/or deductive reasoning strategies (see Table 1.1). The second group of elements includes those which are dependent upon the test constructor or test purpose: intent consideration strategies and/or cue-using strategies (see Table 1.2).

Table 1.1: Taxonomy of Test-Wiseness Principles: Independent of Constructor

---

I. Elements independent of test constructor or test purpose.

(A) Time-using strategy.

1. Begin to work as rapidly as possible with reasonable assurance of accuracy.
2. Set up a schedule for progress through the test.
3. Omit or guess at items (see I. C. and II. B.) which resist a quick answer.
4. Mark omitted items, or items which could use further consideration, to assure easy relocation.
5. Use time remaining after completion of the test to reconsider answers.

(B) Error-avoidance strategy.

1. Pay careful attention to directions, determining clearly the nature of the task and the intended basis for response.
2. Pay careful attention to the items, determining clearly the nature of the question.
3. Ask examiner for clarification when necessary, if it is permitted.
4. Check all answers.

(C) Guessing strategy.

1. Always guess if right answers only are scored.
2. Always guess if the correction for guessing is less severe than a "correction for guessing" formula that gives an expected score of zero for random responding.
3. Always guess even if the usual correction or a more severe penalty for guessing is employed, whenever elimination of options provides sufficient change of profiting.

(D) Deductive reasoning strategy.

1. Eliminate options which are known to be incorrect and choose from among the remaining options.
2. Choose neither or both of two options which employ the correctness of each other.
3. Choose neither or one (but not both) of two statements, one of which, if correct, would imply the incorrectness of the other.
4. Restrict choice to those options which encompass all of two or more given statements known to be correct.
5. Utilize relevant content information in other test items and options.

(From Millman, Bishop, and Ebel, 1965, pp.707 - 726)

Table 1.2: Taxonomy of Test-Wiseness Principles: Dependent Upon Constructor

---

II. Elements dependent upon the test constructor or purpose.

(A) Intent consideration strategy.

1. Interpret and answer questions in view of previous idiosyncratic emphases of the test constructor or in view of the test purpose.
2. Answer items as the test constructor intended.
3. Adopt the level of sophistication that is expected.
4. Consider the relevance of specific detail.

(B) Cue-using strategy.

1. Recognize and make use of any consistent idiosyncrasies of the test constructor which distinguish the correct answer from incorrect options.
  - a. He makes it longer (shorter) than the incorrect options.
  - b. He qualifies it more carefully, or makes it represent a higher degree of generalization.
  - c. He includes more false (true) statements.
  - d. He places it in certain physical positions among the options (such as in the middle).
  - e. He places it in a certain logical position among an ordered set of options (such as the middle of the sequence).
  - f. He includes (does not include) it among similar statements or makes (does not make) it one of a pair of diametrically opposite statements.
  - g. He composes (does not compose) it of familiar or stereotyped phraseology.
  - h. He does not make it grammatically inconsistent with the stem.
2. Consider the relevancy of specific detail when answering a given item.
3. Recognize and make use of specific determiners.
4. Recognize and make use of resemblances between the options and an aspect of the stem.
5. Consider the subject matter and difficulty of neighboring items when interpreting and answering a given item.

(From Millman, Bishop, and Ebel, 1965, pp.707 - 726.)

Since Millman et al.'s seminal work, several other authors have studied and pointed to the deleterious effect of test-wiseness upon the interpretation of test scores (Benson, Urmann, & Hocevar, 1986; Dreisbach & Keogh, 1982; Fueyo, 1977; Kalechstein, Kalechstein, & Docter, 1981; Oakland, 1972; Prell & Prell, 1986; Sarnacki, 1979; Slakter, Koehler, & Hampton, 1970b; and Wahlstrom & Boersma, 1968). Simply, scores of test-wise students exceed those of test-naive students of equal ability (Sarnacki, 1979). As well, this ability, or lack of ability, is not limited to any particular age group. Fueyo (1977) noted that, "[children], as well as adults, can be handicapped when taking a standardized test because of an unfamiliarity with the test format or with the requirements of the testing situation" (p.180). Further, the problem exists both on teacher made tests (Dolly & Williams, 1983b) and commercially available tests (Benson, 1988).

## **1.2 The British Columbia Provincial Examinations**

It is important, at this point, to note the procedures used to construct the provincial examinations. For each content area tested, a subject area committee consisting of teachers is formed. These teachers are selected by an Examination Coordinator and the Manager of Examinations, Student Assessment Branch, Ministry of Education. Each subject area committee is given the responsibility of developing test items using the table of specifications, prepared by the subject area committee, for that subject area. Items from previous examinations and from piloted item banks are considered for inclusion on the test. Additionally, new items may be developed to insure coverage of the curriculum in the subject area tested. The items are then reviewed by measurement specialists, as well as by a different panel of practicing teachers for the subject area before the test is formatted, printed, and distributed. Machine scoreable, selection (multiple choice) items constitute 50% to 80% of the questions included in each test; the remainder of the items

are written supply (subjective) questions.

Following the administration of the tests (January and June for semestered schools, June for non-semestered schools, and August) to all students wishing to receive credit in examinable courses, marking committees comprised of practicing school teachers mark the supply items. The scores for these items, along with the responses to each of the selection items, are then entered into a computer for scoring and item analyses. The results of the item analyses, together with the comments of the marking committee and the technical scoring staff, are then submitted to the Provincial Board of Examiners for its review and consideration. Items are retained or rejected from the students' final scores on the basis of the decisions made by the Provincial Board of Examiners. Typically, very few items on the examinations are rejected (R. Tomusiak, personal communication, February 16, 1989).

### **1.3 Statement of the Problem**

As found by Rogers and Bateson (1990a) students bring varying levels of test-wiseness to the provincial examinations. In their study, Rogers and Bateson focussed primarily on the effects of this source of variability upon performance on the multiple-choice questions included in the provincial examinations. However, the effect of test-wiseness upon performance on the open-ended (supply) questions of the provincial examinations had yet to be investigated.

The purpose of the present study, therefore, was to address the issue of test-wiseness and its relationship to performance on open-ended exercises (supply items). Sarnacki (1979), in his summary of test-wiseness research, stated test-wiseness "...is not limited to objective tests, nor is it limited to the specific strategies delimited by Millman et al.



(1965)" (p.263). Since his review, the issue of test-wiseness and supply test item performance has yet to be investigated. Thus, two of the research questions to be addressed in the present study were: (Ii) What is the nature and the strength of the relationship between test scores on the supply items on the provincial examination in English and test-wiseness? and (Iii) Do students with test-wiseness skills outperform, both in terms of level and variability of performance, those who do not possess such skills on the supply items?

As the measures of test-wiseness in existence use selection-type items and the presence of test-wiseness needs to be demonstrated, two additional research questions (parallel to those stated previously) were proposed. These were: (IIi) What is the nature and the strength of the relationship between test scores on the selection items on the provincial examination in English and test-wiseness? and (IIii) Do students with test-wiseness skills outperform, both in terms of level and variability of performance, those who do not possess such skills on the selection items?

Test-wiseness is considered to be one of many sources of unwanted variance, or bias, on tests (Benson, 1988; Crehan, Koehler, & Slakter, 1974; Kalechstein, & Docter, 1981; Oakland, 1972; Petty & Harrell, 1977; Prell & Prell, 1986; Sarnacki, 1979; Slakter, Koehler, & Hampton, 1970a; Slakter et al., 1970b), yet it has been rarely examined in relation to differential group variability. Gibb (1964) noted

...that the effects of training on a complex task have operated in a way commonly encountered, namely that in addition to improving the performance of the Trained Group relative to the Untrained Group, the training also accentuated differences between individuals, increasing the dispersion of the distribution of scores (1964, p.52).

However, he did not explicitly test for this change in variability, choosing instead to

suggest that further examination of student test score variance will lead to better understanding of how test-wiseness affects student test scores. It has been stated that “presence of test-wiseness cues on a test ... differentially reward high test-wise students while at the same time penalizing students low in test-wiseness” (Prell and Prell, 1986, p.7). It has also been suggested that if the effects of test-wiseness were somehow reduced or eliminated, variability on tests due to this factor would be reduced or eliminated. The intent of the second question, in each pair of questions stated above, is to address the validity of this suggestion through the investigation of variability of performance in test-wise versus test-naive groups.

It has been noted that test-wiseness is correlated with verbal ability (Ardiff, 1965; Benson, 1988; Diamond & Evans, 1972; Rowley, 1974; Sarnacki, 1979). For this reason the relationship between performance on the supply items and test-wiseness will be examined by first controlling for verbal ability.

#### 1.4 Definitions Used In This Study

For purposes of this study the following definitions have been adopted:

**Test-wiseness:** (also called test-taking strategies, test-sophistication, test-familiarization, test-taking orientation, and test-wisdom) “a subject’s capacity to utilize characteristics and formats of the test and/or test taking situation to receive a high score. test-wiseness is logically independent of the examinee’s knowledge of the subject matter for which the items are supposedly measures” (Millman et al., 1965, p.707). The definition of test-wiseness will be limited to those strategies measured by the Test of Test-Wiseness (TTW): absurd options (ID1), similar options (ID2), opposite options (ID3), and stem option similarity (IIB4).

**Test-Wise:** students receiving a score of 18 or above on the TTW (Rogers & Bateson, 1990a).

**Test-Naive:** students receiving a score of 10 or below on the TTW (Rogers & Bateson, 1990a).

**Supply Item:** (also called open-ended, completion, essay, subjective, free response, and restricted free response item) "questions requiring the student to respond in writing. ... may require relatively brief responses or extensive responses" (Sax, 1974, p.578). Supply items were further divided into short-answer and extended-answer items.

**Selection Item:** (also called multiple-choice, objective, and structured response item) "one in which the examinee must choose his answer from the options supplied by the test maker rather than producing it himself. This type of item is called an objective item. The form of the objective items may be alternate-response such as true-false, multiple-choice with three to five answer choices, or matching" (Thorndike & Hagen, 1969, p.93).

### 1.5 Selection of the Dependent Variable

As previous studies in the areas of test-wiseness have pointed to a correlation between verbal ability and test-wiseness (Ardiff, 1965; Diamond & Evans, 1972) and larger gain scores on tests of verbal ability (French & Dear, 1959; Frankel, 1960; Pallone, 1961), it was felt that the English 12 examination would prove to be susceptible to skills in test-wiseness. Analysis of the multiple-choice items on the Provincial Examinations, including the English 12 examinations for 1988, revealed that there were such cues on the January and June forms of the English 12 examination. Examination of the English 12 examination (June, 1989) revealed 21 test-wise susceptible items out of the 28 asked

(Rogers & Bateson, 1990a). Further, all students in British Columbia wishing to graduate from high school must write either the English 12 or Communications 12 provincial examinations. Since the majority of students, 21 338 of 32 306 (A. Friske, personal communication, May, 1989) elect to write the English 12 examination, the sample of students will be representative of the "typical grade twelve student". The sample is not restricted to students with abilities in the sciences or humanities, which would have been the case had any other subject area examination (eg., Chemistry, History) been chosen.

### 1.6 Hypotheses of the Study

It was hypothesized that performance on the supply items of the English 12 component of the British Columbia Provincial Examinations is positively and linearly related to test-wiseness. Verbal ability will be controlled for as it has been found to correlate with test-wiseness in various studies (Ardiff, 1965; Benson, 1988; Diamond & Evans, 1972; Sarnacki, 1979).

$$H_0: \rho_{E(T.V)} = 0$$

$$H_1: \rho_{E(T.V)} \neq 0$$

where  $\rho_{E(T.V)}$  was the semi-partial correlation between performance on the supply items of the June, 1989 English 12 examination (E), and test-wiseness as measured by the TTW (T), first accounting for verbal ability as measured by the LPI (V).

It was further hypothesized that the scores on the supply items included in the English 12 examination of the British Columbia Provincial Examinations of students who

possess test-wiseness skills and abilities would be a) higher but b) less variable than the scores of students who do not possess these same abilities, controlling for verbal ability.

More specifically, the corresponding statistical hypotheses were:

$$\text{Iii (a): } H_0 \quad \mu_{TW} - \mu_{TN} = 0$$

$$H_0 \quad \mu_{TW} - \mu_{TN} > 0$$

where  $\mu_{TW}$  was the mean score on the English 12 examination (supply items) of test-wise students adjusted for verbal ability and  $\mu_{TN}$  was the mean score on the English 12 examination (supply items) of test-naïve students adjusted for verbal ability.

$$\text{(b): } H_0 \quad \frac{\sigma_{TW}^2}{\sigma_{TN}^2} = 1$$

$$H_1 \quad \frac{\sigma_{TW}^2}{\sigma_{TN}^2} > 1$$

where  $\sigma_{TW}^2$  was the test score variance on the English 12 examination (supply items) of test-wise students adjusted for verbal ability and  $\sigma_{TN}^2$  was the test score variance on the English 12 examination (supply items) of nontest-wise students adjusted for verbal ability.

As an anchor for the study, or a basis upon which to judge the relative test-wiseness of the students participating in the study, the performance on the multiple-choice (selection) items of the English 12 examination of the British Columbia Provincial Examinations was investigated. The corresponding hypotheses were:

$$\text{Iii: } H_0 \quad \rho_{E(T.V)_{sel}} = 0$$

$$H_1 \quad \rho_{E(T.V)_{sel}} \neq 0$$

where  $\rho_{E(T.V)_{adj}}$  was the semi-partial correlation between performance on the selection items of the June, 1989 English 12 examination (E), and test-wiseness as measured by the TTW (T), first accounting for verbal ability as measured by the LPI (V).

It was further hypothesized that the scores on the selection items included in the English 12 examination of the British Columbia Provincial Examinations of students who possess test-wiseness skills and abilities would be a) higher but b) less variable than the scores of students who do not possess these same abilities, controlling for verbal ability.

More specifically, the corresponding statistical hypotheses were:

$$\text{Hii (a): Hi } \mu_{TW} - \mu_{TN} = 0$$

$$\text{Ho } \mu_{TW} - \mu_{TN} > 0$$

where  $\mu_{TW}$  was the mean score on the English 12 examination (selection items) of test-wise students adjusted for verbal ability and  $\mu_{TN}$  was the mean score on the English 12 examination (selection items) of test-naive students adjusted for verbal ability.

$$\text{(b): Ho } \frac{\sigma_{TW}^2}{\sigma_{TN}^2} = 1$$

$$\text{Hi } \frac{\sigma_{TW}^2}{\sigma_{TN}^2} > 1$$

where  $\sigma_{TW}^2$  was the test score variance on the English 12 examination (selection items) of test-wise students adjusted for verbal ability and  $\sigma_{TN}^2$  was the test score variance on the English 12 examination (selection items) of nontest-wise students adjusted for verbal ability.

## 1.7 Organization of the Thesis

The remainder of this thesis will be divided into four chapters. Chapter 2 is a review of relevant research literature in the areas of test-wiseness and the British Columbia Provincial Examinations. Chapter 3 will describe the research design and data collection procedures used to collect data necessary to test hypotheses presented above. The results of the analyses of these data and the test of the hypotheses will be presented in Chapter 4. The final chapter, Chapter 5, includes a summary of the study and its results, the limitations of the study, the conclusions that can be drawn in light of these limitations and, lastly, implications for practice and future research.

## **Chapter 2**

### **Review of the Literature**

The review of the literature presented in this chapter has been divided into three sections. Within the first section the definition and nature of test-wiseness is reviewed. Research on correlational, developmental, and variability aspects of test-wiseness, as well as the relationship between supply items and test-wiseness is presented in the second section. In the third section the English 12 examination is examined.

#### **2.1 Test-Wiseness**

##### **2.1.1 Definition(s) of Test-Wiseness**

Thorndike (1951) is considered to be the first to explicitly recognize test-wiseness as an additional source of variance in test scores. In addition to the influence upon scores and the ability to comprehend and follow directions, a

test score is likely to be in some measure a function of the extent to which an individual is at home with tests and has a certain amount of sagacity with regard to tricks in taking them. Freedom from emotional tension, shrewdness with regard to when to guess, and a keen eye for secondary and extraneous cues are likely to be useful in a wide range of tests, particularly those which are not well constructed. The presence of variation in score due to variation in comprehension of instructions and in test-wiseness is understandably undesirable from the point of view of the purpose of the test in question. It usually



represents systematic invalid variance serving systematically to reduce the validity of the test. However, these factors must be recognized. They present a challenge to the author of the test, who will try to minimize them, by providing the clearest possible instruction and a minimum of secondary cues. These factors present a problem of validity rather than one of reliability; as far as our present analysis is concerned, they represent a general, lasting quality of the individual and must be treated as such (pp.568-569).

Gibb (1964), noting that no empirical studies of test-wiseness had been reported since Thorndike's introduction of test-wiseness, defined test-wiseness "tentatively as the ability to react profitably to the presence of secondary cues in a test" (p.5). He limited his definition to "secondary cues ... on a multiple-choice test of knowledge of factual information" (p.5).

Following Gibb, Millman, Bishop, and Ebel (1965) proposed that test-wiseness was "a subject's capacity to utilize characteristics and formats of the test and/or test taking situation to receive a high score. Test-wiseness is logically independent of the examinee's knowledge of the subject matter for which the items are supposedly measures" (p.707). This definition has come to be the most frequently cited definition for test-wiseness. To clarify their definition Millman et al. developed a two part taxonomy of test-wiseness principles. Part 1 contained elements/skills dependent on the test-taker. Part 2 consisted of elements dependent on the test constructor or test items (see Table 1.1 and Table 1.2, Chapter 1).

With relation to elements dependent on the test taker, the skills are organized in terms of five "test taking" strategies: time-using strategies, error-avoidance strategies, guessing strategies, and deductive reasoning strategies. The time-using strategies refer to pacing of responding, omitting items, and returning to omitted items if time allows. The

error-avoidance strategies resemble what Thorndike (1951) had referred to as comprehension of instructions. Emphasis is placed on paying attention to the information present in the directions and within the items themselves, as well as asking for clarification when necessary and reviewing answers. With respect to guessing, the strategy described suggests guessing is appropriate except where a severe penalty for guessing is applied and the test-taker cannot sagely eliminate any of the options present. Deductive reasoning strategies include the elimination of incorrect or absurd options, choosing both or neither of options that are similar, choosing one or neither of two diametrically opposed options, choosing options which include at least two (or more) statements which are known to be true, and using information present in other test items to aid in answering the test item under consideration.

With relation to elements dependent upon the test constructor or test items themselves, Millman et al. (1965) outlined two strategies. These were intent consideration strategies and cue-using strategies. Intent consideration strategies involve the examinee considering the method by which the test was constructed. Additionally the test taker is to consider the intent of the test constructor and past experiences in dealing with the same test constructor. Cue-using strategies include recognition of test constructor idiosyncrasies such as use of longer (or shorter) correct alternatives, greater qualification or generalization of the correct response, inclusion of more true (or false) statements, placing the correct option in one position (such as "c") more often, the use of similar options, the use of stereotypical phrases, and the presence of grammatical cues. Elements under the heading of "Cue-using Strategy" include the use of specific determiners such as all or none, resemblances between the stem and the correct option, the use of relevant (or irrelevant) details, and the difficulty, as well as the subject matter, of the items surrounding the item in question (Millman et al., 1965, pp.707 – 726).

The Taxonomy was created by administering an unstructured questionnaire to 240 high-achieving suburban high school students who were asked to describe their test taking behaviours. Millman et al. found that many of the students were capable of verbalizing several test-wiseness principles. Additional test-wiseness elements included in the Taxonomy were identified from a review of the literature related to test construction.

The Taxonomy of Test-wiseness Principles has served as the basis for subsequent definitions of test-wiseness. Oakland (1972) employed the Taxonomy and defined test-wiseness as “the ability to manifest test-taking skills which utilize the characteristics and format of a test and/or test-taking situation in order to receive a score commensurate with the abilities being measured” (p.355). Diamond and Evans (1972), in their research, defined test-wiseness as “the ability to respond advantageously to multiple choice items containing extraneous clues and to obtain credit on these items without knowledge of the subject matter” (p.145). Sarnacki (1979) pointed out

[test taking] skills in general, and TW [test-wiseness] in particular, generalize to a number of situations. These composite faculties allow examinees not only to exploit specific test flaws, but also to experience a general sense of security in taking tests. TW then, is not limited to objective tests, nor is it limited to the specific strategies delimited by Millman et al. (1965) (p.263).

Williams and Dolly (1983a) considered test-wiseness to be the “ability of the test-taker to perform at a better than chance level on a multiple choice test no matter what the content being tested” (p.2). Rogers and Bateson (1988) more recently defined test-wiseness as a

cognitive ability or set of skills which a test taker can use to improve a test score no matter what the content of the test. If a test taker possesses test-wiseness, and if the examination contains susceptible items, then the combination of these two factors can result in an improved score; in contrast, a student low in test-wiseness will tend to be penalized every time he or she takes a test which includes test-wise components (p.1).

As the various definitions reviewed here are based upon, and include components of, Millman et al.'s (1965) definition, it was chosen for the purposes of the present study. As discussed in the next chapter, the following components of test-wiseness were included in the test-wiseness measure: absurd options (ID1), similar options (ID2), opposite options (ID3) , and stem option association (IIB4) (see Table 1.1 and Table 1.2).

### **2.1.2 Test-Wiseness and Bias**

Test-wiseness is considered to be a source of bias – a systematic component of measurement error (Millman et al., 1965; Oakland, 1972; Petty & Harrell, 1977; Prell & Prell, 1986; Sarnacki, 1979; Slakter et al., 1970b; Thorndike, 1951; Wahlstrom & Boersma, 1968). As initially described in the Taxonomy, test-wiseness is most often regarded as “[encompassing] both the method of measurement and the characteristics of the test-taker” (Sarnacki, 1979, p.267). Smith (1982) notes “when the interaction between the characteristics of an item and a person is viewed from the perspective of the item, it is generally considered to be an issue of item construction; when viewed from the perspective of the person, it is considered to be an issue of testwiseness” (p.211). As such test-wiseness is considered to be both a source of bias, when viewed from the perspective of the test taker, and as a source of error variance (usually referred to as systematic error variance), when viewed from the perspective of the test item. However, regardless of the

viewpoint taken, the effect is the same: bias exists.

Each time a group of students writes a test, the test scores of a subgroup may be higher due to possession of test-wiseness. For these students, their test scores are a composite of both their knowledge of the subject matter being tested and the amount of test-wiseness they possess. For these students test-wiseness is a source of bias, test-wiseness has yielded test scores which reflect an inflated indication of their actual knowledge of what is ostensibly being tested. In contrast, the test scores of nontest-wise students will not be inflated. An inequity exists. Urman (cited in Prell & Prell, 1986) suggests that "a lack of test-wiseness can penalize certain students and the bias against students who are not test-wise ..." (p.3). As well "if differences due to a lack of test-wiseness could be reduced, then error variance attributed to test-wiseness could be minimized" (Prell & Prell, 1986, p.3).

### 2.1.3 The Nature of Test-Wiseness

Thorndike (1951) suggested that test-wiseness is (1) a lasting general trait in relation to an individual's ability to guess strategically and identify cues and (2) a lasting specific trait which is associated with certain types of tests and item formats (pp.568 - 569). Stanley (1971), continued with Thorndike's earlier description. He considered test-wiseness to be a persistent (lasting) trait of the individual and suggested that it was closely related to basic intellectual ability. He felt that test-wiseness would "likely to enter into any test score, whether [it was] wanted or not" (p.364).

Millman (1966), too, suggested test-wiseness is a pervasive skill. Based on findings obtained when validating his test of test-wiseness he reported that the "construct validation of the separate subscales [absurd options, opposite options, similar options, stem options, and guessing] did not hold up with the exception of a multitrait-multimethod study which indicated that subject matter was only slightly influenced by test-wiseness"

(cited in Benson, 1988, p.8).

Diamond and Evans (1972), as a part of their study of test-wiseness, compared correlations between five subscales (longer correct alternatives, stem-option association, specific determiners, grammatical cues, and overlapping distractors). Working with a sample of 95 sixth-grade children they reported correlations of 0.02 to 0.33 suggesting that the use of secondary cues are actually a set of several specific skills, not one general skill. Use of factor analysis seemed to confirm this notion as the separate subscales did not load together.

More recently, Benson (1985) surveyed fourth- through sixth-grade students ( $n=208$ ) on knowledge of test-wiseness using a 20-item Likert-type scale originally designed to measure use of time, error avoidance strategies, use of cues, and guessing strategies, as well as motivation. Five items were negatively correlated with total test score and were therefore deleted. The 15-item scale ( $rel=0.54$ ) was then subjected to exploratory factor analysis. This, and a further confirmatory factor analysis, produced a four factor solution. These four factors were tentatively labelled: thoroughness, preparation, achievement motivation, and perseverance. Correlations among these four factors ranged from  $-0.08$  to  $0.57$  suggesting that these test-wiseness skills were separable. Further, Benson indicated that the alpha coefficient "was not high enough to be indicative of a unidimensional scale" (1988, p.15).

Benson (1988) suggested that "test-wiseness appears to be composed of multiple dimensions [several specific traits] ... which can be reliably measured in both children and adults" (p.15). At this time research, albeit limited, seems to verify Benson's notion that test-wiseness is a collection of several specific traits (or sub-traits) which are not specific to item content. In simplest terms, it appears that test-wiseness is a multidimensional construct, specific to item format, general in relation to academic subject, and lasting or persistent in nature (Slakter, Koehler, & Hampton, 1970a; Crehan, Gross, & Slakter,

1978; Benson, Urman, & Hocevar, 1986).

## 2.2 Test-Wiseness Research

The review of research into test-wiseness can be conveniently divided into four subsections (i) correlates of test-wiseness, (ii) developmental aspects of test-wiseness, (iii) variability of test-wiseness ability, and (iv) test-wiseness and supply items.

### 2.2.1 Correlates of Test-Wiseness

Research into possible correlates of test-wiseness began with Ardif (1965). For her master's thesis, she constructed a test-wiseness measure intended for use with two groups of third-grade students ( $n=44$ ) and two groups of sixth-grade students ( $n=48$ ). This content free measure was designed to assess carefulness, guessing, and reasoning abilities which were thought to be components of test-wiseness. She then correlated scores from her measure of test-wiseness with scores from an intelligence test (the Otis Quick Scoring Mental Ability Test (Form Alpha B) for grade 3 and the California Short Form Test of Mental Maturity (1957, S-form) for grade 6) and standardized reading tests (the Gates Advanced Primary Reading Test (1958) for grade 3 and the Stanford Achievement Test - Intermediate Reading Subtest (1953) for grade 6). Working with correlations corrected for attenuation, Ardif concluded that reading ability ( $r = 0.85$ , group I;  $r = 0.61$ , group II) and intelligence ( $r = 0.51$ , group I;  $r = 0.76$ , group II) were related to test-wiseness at the third-grade level but not at the sixth-grade level ( $r = 0.10$ , group III;  $r = -0.09$ , group IV;  $r = -0.01$ , group III;  $r = -0.15$ , group IV, respectively).

Diamond and Evans (1972) examined the relationship between test-wiseness and intelligence (Lorge-Thorndike Intelligence Test (Form A)) and test-wiseness and achievement

(Iowa Test of Basic Skills – Achievement Battery (Form 2)). They worked with sixth-grade students ( $n=95$ ) from a suburban Philadelphia school district who had “little, if any, training in test taking strategies” (Diamond & Evans, 1972, p.46). Their measure of test-wiseness was comprised of five 6-item subscales: stem-option association, specific determiners, longer correct alternatives, grammatical cues, and overlapping distractors (see Table 1.1 and Table 1.2). Diamond and Evans (1972) concluded that test-wiseness was related to “some general skill or ability as measured by conventional IQ instruments and achievement test scores [yet was also] quite specific to the particular clue or cue under investigation” (p.149).

Rowley (1974), working with 198 ninth-grade students from a southern Ontario high school, tested students on achievement motivation (Russell, 1969), test anxiety (Achievement Anxiety Test (Alpert and Harber, 1960)), test-wiseness (Slakter et al., 1970), and risk taking (modification of Swinford’s (1938, 1941) method). He also administered tests in mathematics (Canadian New Achievement Test in Mathematics (OISE, 1965)) and vocabulary (Dominion Group Achievement Test: Test 1 (Niagara Edition, Ontario College of Education)), first in a free response (supply) format, then in a multiple choice (selection) format 5 weeks later. Using partial correlations, corrected for the unreliability of the free response score and tested for significance using Lord’s Statistic, he found that while the partial correlations of vocabulary with test-wiseness and risk taking were significant, the same was not true for mathematics. Rowley (1974) concluded:

[i] the use of multiple choice tests can produce scores which favour certain types of examinees and penalize others for reasons not explainable in terms of their knowledge of the material being tested ... [and] [ii] differences between the findings for the mathematics and vocabulary tests suggest strongly that the nature of the material being tested is an important factor in determining



the characteristics or suitability of any test format (p.21).

### 2.2.2 Developmental Nature of Test-Wiseness

Research into the developmental nature of test-wiseness began with a cross-sectional study by Slakter, Koehler, and Hampton (1970a). They examined the growth of four selected elements of test-wiseness (stem-options, absurd options, similar options, and specific determiners) across 7 grade levels (5 through 11). Each element was measured using four items, with the 16 test-wiseness items embedded in 28 legitimate test items designed to measure achievement at appropriate grade levels. Subjects were chosen from two school systems: in New York state, 522 males and 548 females, and in Michigan, 600 males and 691 females. Results from both sites indicated that there were no gender and gender by age effects; however, at both sites there was a linear trend in the data: "...the fifth grade students were able to exhibit stem-option and absurd-options behaviors, [although] similar options attainment did not appear frequently until eighth grade and specific determiners attainment did not appear until ninth grade" (Slakter et al., 1970a, p.121). However, Slakter et al. felt they could not rule out the possibility that students who would be considered to be lower achievers had dropped out of school, leaving only high achieving students in the upper grades as subjects in their first study.

To rule out this possible alternate interpretation, Crehan, Koehler and Slakter (1974) in a second, related study tested 1,049 subjects across 7 grade levels (5 through 11) from New York state. Longitudinal data from 539 subjects who were measured twice, once in 1968 and again in 1970, were available in addition to unmatched longitudinal data and cross-sectional data from the 1970 study. The same measure of test-wiseness developed by Slakter et al. (1970a) was used. Students participating in the study were not told the true nature of the test they were taking. Rather, they were led to believe that they were taking an aptitude test. The findings support Slakter et al.'s (1970a) conclusions:

significant increases in test-wiseness were observed over grades 5 – 8, but no further growth or loss was found beyond grade 8.

Crehan, Gross, and Slakter (1978), in a third study, observed 288 of a possible 391 subjects from the original 1970 study after a further two-year interval. The students, now in grades 9 – 12, were measured using the instruments used in the 1970 study. A sex by year multivariate analysis of variance indicated that “TW [test-wiseness] increases with grade over the interval [8 years] studied ... [and noted that] large individual differences ... persist into the high school grades” (Crehan et al., 1978, p.43).

### 2.2.3 Variability of Test-Wiseness Ability

Other than the observation by Crehan, et al. (1978) that large individual differences in test-wiseness exist among students, variability in performance between individuals has seldom been addressed. Although standard deviations have been reported in many studies (Bangret-Drowns et al., 1983; Callenbach, 1973; Crehan et al., 1978; Dreisbach & Keogh, 1981; Moore et al., 1966; Omvig, 1971; Powers & Alderman, 1983; Swinton & Powers, 1983) no analysis of the variability was reported. In one study, Evans and Pike (1973) discussed their results in relation to the standard deviation of a normed test, the Scholastic Aptitude Test (SAT), but did not study individual differences in performances.

Studies exploring the variability of test-wiseness in relation to training or ability level (Gibb, 1964; Slakter et al., 1970b) have shown

that the effects of training on a complex task have operated in a way commonly encountered, namely that in addition to improving the performance of the Trained Group relative to the Untrained, the training also accentuated the differences between individuals, increasing the dispersion of the distribution of scores (Gibb, 1964, p.52).

### **2.2.4 The Effect of Test-wiseness On Supply or Open-ended Test Items**

The research summarized in the previous sections was focussed upon the influence of test-wiseness upon performance on selection or multiple-choice items. In contrast, research to date has not addressed the relationship between test-wiseness and performance on supply or open ended items. However, Sarnacki (1979) suggested test-wiseness "...is not limited to objective tests, nor is it limited to the specific strategies delimited by Millman et al. (1965)" (p.263). The purpose of the present study, therefore, was to: (i) examine the nature and the strength of the relationship between test-wiseness and performance on open-ended items and (ii) determine whether students with test-wiseness skills outperform, both in terms of level and variability of performance, students who do not possess these skills on open-ended items.

### **2.3 British Columbia Provincial Examination Program**

The first provincial examination was administered in March of 1876, when the British Columbia government formally constituted a system of high schools in the province (Bateson, 1984). In order to enter a high school at that time students had to write an examination designed to measure knowledge of subjects prerequisite to high school studies; these being arithmetic, grammar, spelling, and geography. Then, when graduating from high school, students were required to take a set of final examinations, covering four subject areas, designed to assess knowledge of the subject areas necessary for completion of high school and possible entry into an institution of higher education. Entrance to and completion of high school were dependent upon the grades obtained on these provincial examinations. Those involved with curriculum and program evaluation at the time were also concerned with the standards of teaching and the qualifications of teachers in the province (Bateson, 1984). It was believed that teachers were not able to make

sound decisions about their students' abilities with regard to high school entrance and graduation.

Putnam and Weir (1925), in their survey of the school system for the British Columbia government, noted that the standards of teaching, as well as the qualifications of teachers, had improved over the years. Additionally, they suggested that entrance examinations for high school had several adverse effects which included overemphasis of the subjects to be examined at the expense of the subjects not examined, tendency to teach to the test rather than teaching the full curricula, discouraging weaker students from continuing with their studies, and using test results to evaluate the abilities of the teachers. They also suggested that students were placed under an undue amount of pressure to pass these examinations in order to enter into high school. Though Putnam and Weir were not opposed to examinations they saw the examination system in British Columbia as making school a "mere knowledge factory, where drill upon dull, lifeless subject matter is made an end in itself" (1925, p.262).

Beginning in 1931 provincial examinations were only administered to twelveth-grade students; promotion of students in the lower grades was determined solely by the classroom teacher. In 1937 the requirement that all grade 12 students write provincial examinations was modified; those students attending an accredited high school and who had obtained a C+ standing or better were recommended for graduation without writing provincial examinations. Additionally, achievement tests were developed and administered at other grade levels. This system continued with only minor changes until 1974 when provincial examinations were discontinued. It was felt that such testing was no longer necessary for promotion of students, though it was recognized that student achievement needed to be monitored in some manner (Learning Assessment Branch, 1973).

Between 1976 to 1980 province-wide assessments were introduced to monitor the

accountability of the school system through two programs; the Provincial Learning Assessment Program (1976 – 1977) and the Classroom Achievement Test Program (1980). Initially, program level assessments were in three subject areas: Mathematics, Reading, and Science. These first assessments were administered to students in grades 4, 8, and 12 and were to be conducted every four years. By 1983, thirty-four standardized tests, designed to assess individual students in various subject and at various grade levels had been produced as by-products of the Classroom Achievement Testing Program. These tests were being used by extensively by teachers with 750,000 copies being ordered in 1983 (Bateson, 1984).

Despite the existence of various assessment and testing programs, the general public and politicians questioned the quality of public education in the province. It was felt that the standards were declining and that the remedy was seen to be the reintroduction of provincial examinations. Beginning in January of 1984, provincial examinations (in 13 examinable subjects) for all grade twelve students were re-instituted

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

The test results counted for 50% of the student's final grade with the remaining 50% being determined by the classroom teacher. Beginning in January of 1989 the contribution of provincial test scores to a student's final grade was reduced to 40% (in 15 examinable subjects) as a result of the recommendations of the 1988 Royal Commission on Education.

## **2.4 Grade Twelve Provincial Examinations: English 12 Examination**

The provincial examinations are prepared by committees of practicing teachers having expertise in each subject area and follow the procedures are briefly summarized in Chapter 1 (see pp.5 – 6). As shown in Table 2.3, for the English 12 examination (June, 1989), 28 selection items and 11 supply items: 7 short-answer; 3 extended-answer; and one extended composition, were prepared for five topic areas.

Table 2.3: English 12 Examination, June 1989

Part	Topic	Selection Items		Supply Items		Suggested Time(min)
		No.	Marks	No.	Marks	
A	Editing Skills	10	10			15
B	Reading Comprehension	8	8	4	10	35
C	Poetry	5	5	3	14	30
D	Prose	5	5	3	24	45
E	Composition			1	24	55
	T O T A L S	28	28	11	72	180

The selection items within the section on Editing Skills asked students to choose the grammatically correct form of a word, or words, to be inserted into a sentence. For the section on Reading Comprehension, students were first asked to read the first three pages of a *Readings Booklet* and then to select the best answer (from among four options) to questions posed about the reading selection. The same reading/question format was used as well for Poetry and Prose. However, the selections to be read were different; for Poetry a short poem was presented, while for Prose, a three page short story was used.

The supply items in Part B asked students to provide short answers to four questions about a three page selection entitled *Increasing Life Expectancy Means Growing Role for Euthanasia*. In Part C students supplied short answers to two questions and a longer answer to a third question concerning a poem by L. Cohen entitled *A Kite is a Victim*. In Part D students were asked to supply longer answers to two questions posed on the short story, *The Kool-Aid Wino*. Part E consisted of a 300 – 500 word composition on one of three topics provided. Appendix A contains a copy of the English 12 examination.

#### 2.4.1 Scoring: The English 12 Examination

The selection items were machine scored. The supply items were marked by teams of markers, comprised of high school English teachers who were paid for their work. These

teachers were selected by Ministry of Education officials from among those who volunteered. For supply questions other than the English composition, markers used protocols developed by the committee responsible for preparing the English 12 examination. The scoring guide for the June 1989 English Twelve Composition (see Appendix B for a summary) was developed using the compositions for the June 1988 English 12 examination.

Prior to scoring the compositions, each of the markers (over 150 markers for the June 1989 exam) participated in a training session. The six-point holistic scale was introduced to the markers along with six "anchor papers" chosen to exemplify each of the six points, from unacceptable to excellent, on the marking scale. The markers worked together as a group until there was agreement in marking. Marks given were discussed and justification was provided (M. Kozlow, personal communication, August 14, 1989). To increase marker reliability, markers were asked to conform to the standards established by the anchor papers and the six-point scale. Markers were asked not to focus on individual parts of the composition such as grammar, usage, vocabulary, or spelling. They were made aware of certain biases (such as handwriting, margins, double-spacing, or length) affect scoring. Additionally, they were provided with certain rules for holistic marking (see Appendix C).

Once markers had reached agreement and "standards [were] established" (Ministry of Education, 1988, p.3) the markers proceeded to grade the compositions independently. Each composition was read by two markers. Scores were recorded, by each marker, on separate control sheets. A head marker then compared the scores. If the scores were the same, or contiguous, they were added together and doubled to result in a total possible score of 24. If the scores of the two markers were discrepant by more than one mark the essay was read a third time. At this point "the head marker determine[d] the true score" (Ministry of Education, 1988, p.3). Rereads were possible if requested by the student.

The marking required 11 days for the short-answer and extended-answer items. The



selection items were machined-scored as the Ministry received the tests from the schools (A. Friske, personal communication, May, 1990). The marking was completed in a central location.

## **Chapter 3**

### **Methodology**

The purpose of this study was to examine the effects of test-wiseness upon performance on the English 12 examination. Previous research, summarized in Chapter 2, suggested that standardized tests like the provincial examinations may be susceptible to test-wiseness skills possessed by some students.

Described in this chapter are the procedures which were employed in completing the study. First, a description of the sample is provided followed by a description of the tests administered, the test administration schedule and procedures, and the processes for scoring and data entry. The chapter concludes with a delineation of the statistical analyses performed on the data gathered to test the hypotheses stated initially in Chapter 1. These hypotheses are restated in this chapter for completeness.

#### **3.1 Sample**

The data for the study were obtained from the data collected by Rogers and Bateson (1990a) in their study of the nature and impact of test-wiseness upon provincial examination performance of grade twelve students. Eight-hundred students in nine schools located throughout British Columbia (three in the lower mainland; three – 2 urban, 1 rural – in central British Columbia; and three – 1 urban, 2 rural – in the northern coastal area of British Columbia) were included in their total edited file.

The school districts included in the study were selected so as to provide a school sample representation with consideration for the full range of academic ability, as well as

diversity of socio-economic and ethnic backgrounds and a full set of background variables (Rogers & Bateson, 1990a). The schools were selected in consultation with District Central staff. All selected districts and schools agreed to participate. As well, the Ministry of Education provided Provincial Examination results for 767 of the original sample of 800 students using Provincial Ministry identification numbers to match student files. Of this number, 735 wrote the English 12 examination.

## **3.2 Instrumentation**

### **3.2.1 English 12 Examination**

Performance on the selection and supply items included in the June 1989 form of the English 12 examination formed the dependent variables. Analysis of previous English Twelve Provincial examinations revealed that there were items susceptible to test-wiseness. The 1989 English Twelve test included 28 multiple-choice items and 11 supply items with which to examine the influence of test-wiseness upon such items.

### **3.2.2 Test of Test-Wiseness**

The test-wiseness instrument consisted of 34 questions. Items from instruments developed by Gibb (1964), Millman (1966), and Slakter et al. (1970), some with modification, were used. The instrument was designed to be content free; subject area knowledge was not necessary to answer the questions on the test. In some cases there was no correct answer as the material was entirely fictitious. The instrument was designed to assess the attainment of four different test-wiseness skills identified from an analysis of multiple-choice items included in the previous Provincial Examinations in the areas of English, Geography, History, Biology, Chemistry, and Algebra. A separate section on guessing was also included. The section on guessing was not examined as it was not relevant to

the questions proposed by this study.

In the section of the Test of Test-Wiseness (TTW) used in this study students were asked to select an answer, from among four options, to each question. Students were encouraged to guess if they felt they did not know the answer. Students recorded their responses, using an HB pencil, on separate machine-scoreable answer sheets which were provided.

Table 3.4 contains a description of the final form, a copy of which is included in Appendix D. Administration time was 30 minutes.

Table 3.4: Description of the Test of Test-Wiseness

Test-Wiseness Element	Mathematics	Subject Biology	Area Social Studies	English
<u>Section A</u>				
Absurd option	1 <sup>a</sup> (3) <sup>b</sup>	1(7)	2(13,17)	2(9,22)
Similar options	1(6)	2(12,18)	1(21)	2(4,15)
Different options	2(16,23)	1(1)	1(5)	2(10,20)
Stem-option Link	1(14)	2(2,24)	2(8,19)	1(11)
TOTAL	5	6	6	7
<u>Section B</u>				
Guessing	1(3)	2(5,6)	1(2)	2(9,10)
Non-Guessing	1(8)	1(1)	1(7)	1(4)
TOTAL	2	3	2	3

Note: a(b): a – number of items; b – item number(s)

## Background Information

Background information was collected during the data collection for the test-wiseness instrument. Prior to answering the test-wiseness questions students were asked to provide information regarding their date of birth, grade, gender, ethnic background, previous practice using provincial examinations, and coaching on test-taking skills. Responses

were recorded on the same answer sheet provided for the TTW (see Appendix D for the list of background questions asked).

### **Validation**

To assess the validity of this test, students in six of the schools who scored either above 17 ( $n=36$ ) or below 11 ( $n=41$ ) on part A of the TTW responded to a second test-wiseness test. The composition of this second test is shown in Table 3.5. Unlike the first test, the second test was administered individually. Students were asked to talk aloud as they answered the questions – describing the method by which they arrived at their answers. Strategies used were recorded, as well as any additional information provided by the students. Results indicated that several test-wiseness strategies were being used by the students who had received a score of 17 or more on the 24 questions of the original measure designed to address the use of absurd options (ID1), similar options (ID2), opposite options (ID3), and stem-options (IIB4). While those scoring below 11 did use some of these abilities the incidence of guessing randomly was much greater (Rogers & Bateson, 1990b). Questions used on the validation instrument are presented in Appendix E.

#### **3.2.3 Language Proficiency Index**

Test-wiseness has been found to correlate with measures of verbal ability (Ardiff, 1965; Diamond & Evans, 1972; Benson, 1988; Rowley, 1974; Sarnacki, 1979). Therefore verbal ability, measured by the first two sections of the Language Proficiency Index (LPI (Educational Measurement Research Group: EMRG)), was included to control for this possible source of variation.

The LPI was designed to provide “post secondary schools with a method to determine [a student’s] level of English competence...” (EMRG, 1988, p.1). Using this measure,

Table 3.5: Description of the Test-Wiseness Measure Used for Validation

Test-Wiseness Element	Mathematics	Subject Biology	Area Social Studies
Absurd option	2 <sup>a</sup> (9, 14) <sup>b</sup>		2 (1,8)
Similar options		2 (4,7)	1 (1)
Different options		2 (2,5)	
Stem-option Link	1 (6)	1 (10)	1 (3)
Guessing	3 (9,13,14)	1 (10)	1 (11)
Non-Guessing			1 (12)
TOTAL	6	6	6

Note: a(b): a – number of items; b – item number(s)  
Some questions are found in more than one category.

post secondary institutions were able to determine placement in the most suitable English course for students.

The LPI is composed of two sections. The first section deals with the recognition of common errors in English usage and sentence construction. The second section covers essay composition. Within the first section, multiple-choice items are used to measure performance on the topics of sentence structure, English usage, and the development, structure, and content of paragraphs. Results from this section are to be used as “supplementary placement data” (EMRG, 1988, p.4). The second section is “the composition of a 300 – 400 word expository essay” (EMRG, 1988, p.4) based on the selection of one of a wide range of topics provided. Results from this section are to be used as “the main determinant of placement” (EMRG, 1988, p.4).

Ten questions were selected from an item pool for Sentence Structure and ten from an item pool for English Usage for the form of the LPI used by Rogers and Bateson. A copy of the LPI used in the present study is provided in Appendix F.

### 3.3 Procedures

#### 3.3.1 Testing

The testing schedule followed is shown in Table 3.6.

Table 3.6: Testing Schedule

School	May	May	June
1	TTW	LPI	English 12
2	TTW	LPI	English 12
3	TTW	LPI	English 12
4	TTW	LPI	English 12
5	TTW	LPI	English 12
6	TTW	LPI	English 12
7	LPI	TTW	English 12
8	TTW	LPI	English 12
9	TTW	LPI	English 12

Testing took place within one class period of 30 minutes for the test-wiseness measure. The LPI, which was administered in a separate class period as a part of a test battery, required approximately 15 minutes to complete. It was felt that the LPI would not influence test scores on the TTW, and the TTW would not influence test scores on the LPI (as they were measuring very different constructs), therefore the order of test administration was not the same for all classes participating in the study.

The administration of the TTW was, for the most part, completed by Rogers, Bateson, and the principal investigator of the present study. When the investigators were not able to administer the measures, either due to time or distance constraints, the classroom teacher administered the tests using a protocol for the test (see Appendix G). The LPI was administered by the classroom teacher using the test protocol developed for the LPI (see Appendix H). The English 12 examination was administered by the classroom teacher in accordance with the guidelines set out by the Ministry of Education in British

Columbia.

Instructions given during the test administration included information about the length of time for test administration. Students were asked to answer all questions as best they could for the test-wiseness section of the TTW and the LPI. For the guessing section of the TTW students were informed of a penalty for guessing formula to be used when correcting that section: "If the student did not answer the question they would receive zero. For a correct answer the student would receive five points and two points would be deducted for an incorrect answer." Students were asked to record their answers on the answer sheets provided for the TTW. For the LPI, students were allowed to answer the questions directly on the test provided. Further help was not given during the testing sessions; students were encouraged to proceed as best they could.

### 3.4 Scoring and Data Entry

The TTW was marked, using machine scoreable answer sheets, by the Educational Measurement Research Group (EMRG). Prior to marking, the tests were examined for stray marks and other possible problems. Integrity of this process was further insured by randomly selecting 20 answer sheets and comparing these to the data in the mainframe computer file. The LPI item responses were hand entered and recorded on magnetic tape by a professional data entry organization, ELAN Data Makers, with 100% verification. A further check on the integrity of the data was made by randomly selecting ten tests and comparing the answers to those provided in the mainframe file.

Both the TTW and the LPI were scored using the item analysis package LERTAP (Nelson, 1974) which at the same time yielded estimates of internal consistency. The scores were then merged into a single file for subsequent analyses. Data integrity was insured by randomly checking the identification numbers by which the information had



been sorted. Each test had been assigned a provincial identification number for the student writing that test.

The English 12 examination was scored by the Ministry of Education following the procedures described in Chapter 2. The merged file containing provincial identification numbers and information on the LPI and the TTW were copied onto magnetic tape and sent to the Ministry. This file was merged with test results of the English 12 examination by the Ministry in September of 1989. Provincial policy required that the identification codes be changed by the Ministry to ensure the anonymity of the students participating in the study.

### **3.5 Statistical Analysis**

#### **3.5.1 Preliminary Analyses**

Prior to conducting the statistical analyses to test the hypotheses of this study, a series of preliminary analyses were conducted to test for differences among schools. A oneway analysis of variance was completed separately for each of the test-wiseness elements measured – ID1, ID2, ID3, and IIB4, and the two subtests of the LPI, using the SPSS-X computer program ONEWAY (Norussis/SPSS Inc., 1988). Working at the 0.20 level of significance, no school differences were found (see Appendix I). Consequently, the unit of analysis was set at the student level ( $n=735$ ).

#### **3.5.2 Description of the Sample**

The background information collected (see Section 3.2.2) was analyzed using the SPSS-X computer program FREQUENCIES (Norussis/SPSS Inc., 1988). Frequencies were calculated for each of the levels of gender, ethnicity, and birth year, and coaching and practice in test-taking.

## Hypothesis II

To test the hypothesis that performance on the supply component, for short- and extended-answer questions, of the English 12 examination was linearly and positively related to test-wiseness after taking account of verbal ability, an *a priori* linear regression analysis (Pedhauser, 1984) was completed using SPSS-X REGRESSION.

The model tested was:

$$Y = b_0 + b_V V + b_T T + e$$

where  $Y_{SA}$  = performance on the short-answer items

of the June, 1989, English 12 examination,

$Y_{EA}$  = performance on the extended-answer items

June, 1989, English 12 examination,

$V$  = verbal ability as measured by the LPI,

$T$  = test-wiseness as measured by performance on  
the TTW, and

$b_V$  and  $b_T$  are the regression weights (partial regression coefficients)  
for verbal and test-wiseness ability.

Testing the significance in the increase in  $R^2$  brought about by the inclusion of  $T$  after  $V$  provided a test of significance of the corresponding semi-partial correlations  $\rho_{E(T.V)_{SA}}$  and  $\rho_{E(T.V)_{EA}}$ . The semi-partial correlation coefficient was calculated using the formula:

$$pr_{E(T.V)} = \frac{r_{ET} - r_{EV}(r_{TV})}{\sqrt{1 - r_{TV}^2}}$$

(Glass & Hopkins, 1984, p.130)

Supply items were separated, on the basis of length of response, into short-answer and extended-answer items. It was felt that these item types may not be equivalent due to the nature of the response that was required.

### Hypothesis Iii

Following the same procedure as that used to test hypothesis Ii, the second hypothesis related to the significance of the semi-partial correlation,  $\rho_{E(T.V)_{SEL}}$  where  $Y_{SEL}$  was the performance on the 28 selection items of the English 12 examination (June, 1989), was tested.

### Hypothesis Iii (a and b)

A one-way analysis of covariance (ANCOVA) was used to test the hypotheses of non-significant differences between the mean of students possessing test-wiseness adjusted for verbal ability and the adjusted mean of students who do not possess test-wiseness for the short-answer and extended-answer components of the English 12 examination (Glass & Hopkins, 1984).

The variances associated with the means, adjusted for verbal ability, were tested using the F-test for two independent variances,  $\sigma_{TW}^2$  and  $\sigma_{TN}^2$ , for the short- and extended-answer sections of the English 12 examination (Glass & Hopkins, 1984).

### Hypothesis Iiii (a and b)

Following the same procedures to test hypotheses Iii (a and b), the second set of hypotheses related to the significance of the adjusted means and variances for the test-wise versus test-naive groups where  $\mu_{TW}$ ,  $\mu_{TN}$ ,  $\sigma_{TW}^2$ , and  $\sigma_{TN}^2$  are based upon the scores on the 28 selection items of the English 12 examination, was tested.

## **Chapter 4**

### **Results**

This chapter is divided into two major sections, Substudy I and Substudy II, relating to the hypotheses presented in Chapter 1. Substudy I presents the results for hypotheses Ii and Iii. These hypotheses dealt with the nature and strength of the relationship between test-wiseness ability (as measured by the TTW) and the English ability (as measured by the English 12 examination). Substudy II relates the results for hypotheses IIi (a and b) and IIii (a and b). Analyses of the results with respect to differences in means and variances, adjusted for verbal ability for test-wiseness and test-naïve subjects, are discussed. Results for Substudy I are preceded by descriptions of the sample and instruments used. Results for Substudy II are preceded by a description of the samples used.

#### **4.1 Substudy I**

##### **4.1.1 Description of the Sample**

As reported in the previous chapter, the subjects in the present study were obtained from a data bank obtained by Rogers and Bateson (1990a) in their study of test-wiseness. Altogether, complete data consisting of scores on the multiple-choice, short-answer, and extended-answer components of the English 12 examination, the LPI, and the TTW were available for 735 subjects.

Background information on gender, age, ethnic origin, and prior experience and training in test-taking for the full sample are summarized in Table 4.7.

Table 4.7: Description of the Full Sample

Independent Variable	n	percentage
<b>Gender</b>		
Female	392	53.2%
Male	340	46.4%
<b>Birthdate</b>		
1954	1	0.1%
1967	1	0.1%
1968	4	0.5%
1969	5	0.7%
1970	87	11.8%
1971	614	83.5%
1972	14	1.9%
<b>Ethnic Background</b>		
English	428	58.2%
French	21	2.9%
Native Indian	11	1.5%
East Indian	30	4.1%
Chinese	36	4.9%
German	43	5.9%
Italian	24	3.3%
Japanese	8	1.1%
Other	131	17.8%
<b>Information on Test Preparation</b>		
Coaching		
No, never	399	54.3%
Yes, once or twice	263	35.8%
Yes, three or more times	70	9.5%
Practise		
No, never	131	17.8%
Yes, once or twice	373	50.7%
Yes, three or more times	223	30.3%

Note: Data for the missing cases were not included. It was reflected in the percentages.  
 Coaching - "Have you ever had any coaching or specific lessons on how to take a test?"  
 Practise - "Have you ever practised writing provincial examinations using previous provincial examinations or questions from these examinations?"

As shown in Table 4.7 the number of male and female students for this sample is approximately equal. The majority of students, over 83%, were born in 1971 thus being, or becoming before year end, 18 years of age at the time of data collection. Nearly 60% of the students indicated that they were of English origin, while over 17% indicated that their ethnic origin was other than one of the eight categories provided. While slightly more than 54% of the students reported they had never received information on, or training in, test taking practices or procedures, over 80% of the students, reported they had practice writing at least one or two previous provincial examinations.

#### 4.1.2 Test Characteristics

Statistical analyses of the data, presented later in this chapter, were based upon total test score for the TTW ( $k=24$ ), total test score for the LPI ( $k=20$ ), and test scores for the multiple-choice ( $k=28$ ), short-answer ( $k=14$ ) and extended-answer ( $k=1$ ) items types found on the English 12 examination. Multiple-choice and supply-type items were examined separately as the latter had yet to be addressed in the literature (Sarnacki, 1979).

Table 4.8 provides means, standard deviations, and internal consistencies for the measures used. These data are presented for both the full sample of 735 subjects and the provincial population of 21,338 students. Short and extended-answer items ( $r = 0.93$ ;  $r = 0.69$ ;  $n=735$ , respectively, representing the relationship to the total examination) proved to be dissimilar, therefore separate analyses were conducted.

The means and standard deviations for the full sample (18.40; 3.62) and the population (18.4; 3.71) for the multiple choice items indicated that the sample did not vary significantly ( $t = 0.00$ ;  $F = 1.05$ , respectively) from the population.

Information for the means and standard deviations for the short- and extended-answer items were not provided by the Ministry. Instead, as shown in Table 4.8,

Table 4.8: Means, Standard Deviations, and Internal Consistency of Tests Analyzed for Full Sample and Provincial Data

	$k^a$	Mean		Standard Deviation		Reliability	
English 12 examination							
Multiple-choice	28	18.4 <sup>b</sup>	(18.4) <sup>c</sup>	3.62	(3.71)	0.61	
Short-answer	14	29.5		7.85		0.75	(N/A) <sup>d</sup>
Extended-answer	1	14.9		3.30		N/A	(N/A)
Supply (total)	15	44.4	(45.7)	9.82	(10.49)		
Total Test	43	62.8	(64.1)	12.29	(13.03)	0.79	(0.80)
Test of Test-Wiseness	24	14.0		2.79		0.37	
Language Proficiency Index	20	12.0		3.56		0.69	

Note: a  $k$  = total number of items short-answer and extended-answer items being differentially weighted  
 b (c) where b = full sample ( $n=735$ )  
 c = provincial population ( $N=21,338$ )  
 d (N/A): inter-rater reliability information was not available from the Ministry (A. Chatwynde, May, 1990, personal communication).

the means and standard deviations were available for the sub-score computed from these items considered together as well as for the total test score. While the mean for the sample was significantly different ( $t = 3.28$ ;  $t = 2.64$ ;  $p < 0.01$ , respectively) from that of the population, inspection of the means (44.5 vs. 45.7; 62.8 vs. 64.1) revealed they were quite close. The significance is likely an artifact of the large sample size (hence the degrees of freedom of the t-test). Taken as a whole, the results suggest that the performance of the sample students was representative of that for the population.

Internal consistency estimates (Hoyt, 1941) for the multiple-choice and short-answer items are moderately to strongly correlated (0.61 and 0.75, respectively). The reliability for the total test, 0.79, was comparable to that reported by the Ministry of Education, 0.80.



The psychometric characteristics for the TTW for the full sample are presented in Table 4.8. Although the TTW contains four subscales (ID1, ID2, ID3, IIB4) to ensure content validity, only the results from the total test ( $k=28$ ) were used in the statistical analyses completed to test the hypotheses of this study. The obtained reliability for the TTW,  $rel=0.37$ , is comparable to the measures of test-wiseness developed by Millman (1966). Millman (1966) reported reliabilities of 0.53 (KR-20) for a sample of high school students and 0.38 (KR-20) for a sample of college students. As the populations under consideration in Millman's (1966) study were similar in age to those participating in this study, and items used on the TTW were from Millman's (1966) measures of test-wiseness, in original or revised form, comparison of the TTW to these specific measures was considered relevant.

The mean, standard deviation, and internal consistency found for the LPI (see Table 4.8) compare favourably to those commonly obtained by EMRG (D. Blackmore, personal communication, June, 1989).

#### 4.1.3 Results of the Statistical Analyses

It was hypothesized that performance on the selection, short-answer, and extended-answer components of the English 12 examination would be linearly and positively correlated with the residuals formed from the prediction of test-wiseness from language ability. That is to say, performance on selection, short-answer and extended-answer subtests would be linearly and positively related to test-wiseness after language ability was removed from test-wiseness:

$$H_0: \text{If } \rho_{E(T.V)_{SA}} = 0$$

$$\text{If } \rho_{E(T.V)_{EA}} = 0$$

$$\text{Iii } \rho_{E(T.V)_{SEL}} = 0$$

Table 4.9 contains the zero-order correlations among test-wiseness and language abilities with performance on selection, short-answer and extended-answer subtests for the English 12 examination. Table 4.10 contains the values of the corresponding semi-partial correlations. The results of the tests of these hypotheses are presented in Table 4.11.

Table 4.9: Zero-Order Correlation Matrix (n=735)

	Test of Test-Wiseness	Language Proficiency Index
English 12 examination		
Multiple-choice	0.347	0.527
Short-answer	0.314	0.470
Extended-answer	0.226	0.403
Test of Test-Wiseness		0.327

Table 4.10: Semi-Partial Correlations: English 12 Examination by Test-wiseness (Verbal Ability) (n=735)

Item Type	Semi-Partial Correlation (pr)	Variance Accounted for ( $R^2$ )
Multiple-choice	0.185	0.034
Short-answer	0.170	0.029
Essay	0.100	0.010

Test-wiseness and verbal abilities show low to moderate relationship with the selection, short-answer, and extended-answer subtests of the English 12 examination, as illustrated by the correlation coefficients in Table 4.9. As shown in Table 4.10, the values for the semi-partial correlations between english performance and test-wiseness following the removal of verbal ability are: 0.185 for the selection subtest, 0.170 for the short-answer subtest, and 0.100 for the extended-answer subtest. Each is statistically

Table 4.11: Test of Significance of the Semi-Partial Correlation: English 12 Examination by Test-Wiseness (Verbal Ability) (n=735)

Source of Variation	Proportion of Variation	df	F-Ratio
Multiple-choice			
$R_{E.TV}^2$	0.31152	2	
$R_{E.V}^2$	0.27716	1	
1 - $R_{E.V}^2$	0.03436	732	36.5338
Short-answer			
$R_{E.TV}^2$	0.24977	2	
$R_{E.V}^2$	0.22096	1	
1 - $R_{E.V}^2$	0.02881	732	27.0720
Extended-Answer			
$R_{E.TV}^2$	0.17196	2	
$R_{E.V}^2$	0.16203	1	
1 - $R_{E.V}^2$	0.00993	732	8.6748

Note:  $0.01 F_{1,732} = 4.62$

significant ( $p < 0.01$ ). Approximately four percent of the variance for each of the three English 12 subtests was accounted for by the unique portion of test-wiseness. While not large, these findings suggest that test-wiseness does contribute uniquely to the variability in English performance.

As the LPI used a multiple-choice type of format, test-wiseness ability may have influenced test score on the LPI. Use of LPI test-score as the covariate measure may have lessened the power of the statistical test, for each subtest, in assessing the degree of relationship and the amount of variance accounted for by test-wiseness.

Analyses of the residual scores, standardized residuals on test-wiseness and predicted score on the English 12 examination, for the three subtests revealed essentially linear relationships (see Appendix J). However, there were some instances of extreme outliers ( $\pm 2.5$  standard deviations). As the sample from Substudy I and the population were

found to be similar it is suggested that such outliers would be found in the general population, hence deletion of these outliers would not be justified.

To further clarify the relationship between test-wiseness and performance in English, two subsamples – a test-wise group and test-naive group – were selected with TTW test score as an indicator of test-wiseness ability. These results are reported next in Substudy II.

## 4.2 Substudy II

### 4.2.1 Description of the Test-Wise and Test-Naive Samples

To test the second set of hypotheses, two subsamples, drawn from the full sample of 735 subjects, were formed on the basis of their test-wiseness score. Students with scores greater than 17 on the TTW were considered test-wise, while those scoring below 11 were considered test-naive. The number of subjects in the two subsamples were, respectively, 61 and 76.

Cut-scores for the two samples were based upon clinical observation and measurement characteristics of the TTW. Mean TTW scores for subjects in the test-wise and test-naive groups were one standard deviation apart from the grand mean ( $n=137$ ). Further, observations made while validating the TTW indicated that subjects interviewed who had previously scored above 17 on the TTW used test-wiseness strategies to a greater extent than those who scored below 11. Subjects who scored below 11 on the TTW guessed the answers much more frequently than those who scored above 17 on the TTW (Rogers & Bateson, 1990b).

Background information on gender, age, ethnic origin, and prior experience and training in test-taking for each of the subsamples are summarized in Table 4.12.

Table 4.12: Description of the Test-Wise (n=61) and Test-Naive (n=76) Samples

Independent Variable	n		percentage	
	Test-Wise	Test-Naive	Test-Wise	Test-Naive
<b>Gender</b>				
Female	30	45	49.2%	59.2%
Male	31	31	50.8%	40.8%
<b>Birthdate</b>				
1954	0	0	0.0%	0.0%
1967	0	0	0.0%	0.0%
1968	0	2	0.0%	2.6%
1969	0	1	0.0%	1.3%
1970	4	17	6.6%	22.4%
1971	54	55	88.5%	72.4%
1972	2	1	3.3%	1.3%
<b>Ethnic Background</b>				
English	33	43	54.1%	56.6%
French	2	4	3.3%	5.3%
Native Indian	0	3	0.0%	3.9%
East Indian	1	7	1.3%	9.2%
Chinese	2	3	3.3%	3.9%
German	6	4	9.8%	5.3%
Italian	3	0	4.9%	0.0%
Japanese	0	1	0.0%	1.3%
Other	14	11	14.5%	23.0%
<b>Information on Test Preparation</b>				
<b>Coaching</b>				
No, never	36	45	59.0%	59.2%
Yes, once or twice	23	22	36.1%	30.3%
Yes, three or more times	3	7	4.9%	9.2%
<b>Practise</b>				
No, never	9	22	14.8%	28.9%
Yes, once or twice	28	31	45.9%	40.8%
Yes, three or more times	24	20	39.9%	26.3%

Note: Data for the missing cases were not included. It was reflected in the percentages.  
 Coaching – “Have you ever had any coaching or specific lessons on how to take a test?”  
 Practise – “Have you ever practised writing provincial examinations using previous provincial examinations or questions from these examinations?”

The test-wise sample ( $n=61$ ) had equal proportions of males and females. The majority of subjects in this sample, over 88%, were born in 1971. The majority of subjects, slightly over 54%, indicated that they were English. Slightly more than 14% indicated that they were of ethnic group other than those provided on the background questionnaire. Sixty percent of the subjects had received no information on, or training in, test preparation. Slightly over 36% had had some preparation in this area. Over 85% of the subjects in this group had taken at least one provincial examination for practice prior to the collection of the data.

The test-naive sample ( $n=76$ ) contained more females than males (59.2% vs. 40.8%). Over 72% of the subjects were born in 1971. The majority of subjects, over 56%, considered themselves to be of English background, while 23% considered themselves to be of ethnic background other than one of the eight categories specified on the questionnaire. The majority of subjects, slightly over 59%, had never received information on test preparation. Most of the subjects had had practice writing one or two provincial examinations. A total of 67% had written at least one provincial examination for practice prior to collection of the data used in the present study.

The test-wise and test-naive groups were similar in composition. Chi-square analyses were conducted to compare the two groups on the variables of gender, birthdate, ethnicity, coaching, and practice. As cell sizes for birthdate and ethnicity were inadequate, the students were regrouped. Birthdate was dichotomized; students born before 1971 were considered separately from those born in, and after, 1971. Ethnicity was redefined; French, German, and Italian subjects were regarded as European, Chinese and Japanese subjects as Asian, Native Indian, East Indian, and Other subjects as Others, and English subjects remaining in their original category.

Results of the chi-square analyses indicated that the test-wise and test-naive groups were not significantly different with respect to the variables of gender ( $\chi^2 = 1.37$ ;  $df = 1$ ),

ethnicity ( $\chi^2 = 1.23$ ;  $df = 3$ , Yates correction), coaching ( $\chi^2 = 1.19$ ;  $df = 2$ , Yates correction), and practice ( $\chi^2 = 4.93$ ;  $df = 2$ , Yates correction). The test-wise and test-naive groups were statistically significantly different with respect to the variable of birth year ( $\chi^2 = 8.91$ ;  $df = 1$ ;  $p < 0.01$ ). There were proportionally more subjects born before 1971 in the test-naive group as compared to test-wise group.

#### 4.2.2 Results of the Statistical Analyses

The hypotheses for Substudy II, set forth in Chapter 1, dealt with the differences between the means and variances for the test-wise and test-naive groups for three subtests; selection, short-answer, and extended-answer; of the English 12 examination.

It was hypothesized that performance on the English 12 examination for the test-wise group would be higher than that of their test-naive counterparts for each of the three subtests examined, with adjustments for verbal ability. Table 4.13 presents the results of the analyses of covariance for the selection, short-answer, and extended-answer subtests. The means (adjusted for verbal ability) are presented in Table 4.14.

Results from the analyses (see Table 4.13) revealed that the performance of the test-wise and test-naive groups were significantly different ( $p < 0.01$ ) in favour of the test-wise group on the selection and short-answer subtests of the English 12 examination. However, the test-wise and test-naive groups did not differ significantly on the extended-answer subtest. The means for the test-wise and test-naive groups, adjusted for the presence of verbal ability, were 2.45 score points apart for the selection subtest and 5.59 score points apart for the short-answer subtest. In contrast the means for the test-wise and test-naive groups, adjusted for the presence of verbal ability, were 1.00 score points apart for the extended-answer subtest. Results for the analysis of variance, by subtest, were similar for the selection and short-answer subtests ( $F = 56.33$  and  $F = 41.67$ ,  $p < 0.01$ , respectively). However, the results for the extended-answer subtest proved significant

Table 4.13: Results of the ANCOVA Analyses for the English 12 Examination (Selection, Short-Answer, and Extended-Answer Subtests)

Source of Variation	df	Mean Squares	F-Ratio
Selection			
Covariate (Language Proficiency)	1	607.45	75.39
Between groups	1	151.85	18.85
Within groups	134	8.06	
Total	136	13.56	
Short-Answer			
Covariate (Language Proficiency)	1	2030.38	41.48
Between groups	1	780.32	15.94
Within groups	134	48.94	
Total	136	68.89	
Extended-Answer			
Covariate (Language Proficiency)	1	350.66	45.89
Between groups	1	251.16	3.29
Within groups	134	7.64	
Total	136	10.29	

Note:  $0.01 F_{1,134} = 6.85$

( $F = 21.57$ ,  $p < 0.01$ ). Use of verbal ability as a covariate may have reduced the power of the statistical tests in the previous analyses. Further discussion of overlap between test-wiseness and verbal ability is found in Chapter 5.

Results of the residual analyses indicated that there were no unusual patterns, or correlations, in the data (see Appendix K). For the most part, that standardized residuals were within normal limits ( $\pm 2.5$  standard deviations) thus further adjustments to the data were not necessary.



Table 4.14: Adjusted Means and Standard Deviations: English 12 Examination (Selection and Short-Answer Subtests)

Subtest	Group (n)	Mean (covariate)	Mean (subtest)	Adjusted Mean
Selection	Test-Wise (61)	13.80	20.79	19.880*
	Test-Naive (76)	10.00	16.70	17.43
Short-Answer	Test-Wise (61)	13.80	34.18	32.79*
	Test-Naive (76)	10.00	26.09	27.20
Extended – Answer	Test-wise (61)	13.80	16.23	15.46
	Test-Naive (76)	10.00	13.84	14.14

\*= $p < 0.01$

From these results it was concluded that:

$$H_i: \text{Ili (a)}_{SEL} \mu_{TW} - \mu_{TN} > 0$$

$$\text{Ili (a)}_{SA} \mu_{TW} - \mu_{TN} > 0$$

and

$$H_o: \text{Ili (a)}_{EA} \mu_{TW} - \mu_{TN} = 0$$

were tenable.

Discussion of the significance of these results follows in Chapter 5.

Test-score variance for the test-wise and test-naive subjects was examined for the selection, short-answer, and extended-answer subtests of the English 12 examination.

It was hypothesized that students who possess test-wiseness skills and abilities would have test scores that are less varied than the test scores of students who do not possess these same abilities, controlling for verbal ability.

Results of the analyses indicated that test-wise subjects did not have test scores that differed in variability, following adjustment for verbal ability, from those of their

test-naive peers for the selection, short-answer, and extended-answer subtests of the English 12 examination ( $F = 1.27$ ;  $F = 1.03$ ;  $F = 1.07$ ,  $p < 0.01$ , respectively).

From these results it was concluded that:

$$\text{Ho: IIi (b) }_{SA} \frac{\sigma_{TW}^2}{\sigma_{TN}^2} = 1$$

$$\text{IIi (b) }_{EA} \frac{\sigma_{TW}^2}{\sigma_{TN}^2} = 1$$

$$\text{and IIii (b) }_{SEL} \frac{\sigma_{TW}^2}{\sigma_{TN}^2} = 1$$

were tenable.

## **Chapter 5**

### **Conclusions and Recommendations**

This chapter has been divided into five sections. Within the first section a brief synopsis of the problem and procedure are presented. The second section outlines the results of the study, Substudy I and Substudy II being presented separately. Limitations are discussed in the third section. The fourth section provides further discussion of the problem with reference to previous literature as well as the test-wiseness measure. The fifth section presents recommendations based upon the results of the study and review of previous literature. Implications for practice and recommendations for future research are presented separately.

#### **5.1 Summary of the Problem and Procedure**

The purpose of this study was to examine the relationship between test-wiseness and the open-ended (supply) items of the English 12 examination administered in June of 1989. To provide a reference point for interpreting the findings, a set of parallel questions was developed to assess the relationship between test-wiseness and performance on the selection items of the English 12 examination.

Data were available from a sample of 735 students enrolled in nine schools. Comparisons of the characteristics of these students and their performance on the English 12 examination with the population data and results revealed the sample was representative of the provincial population of students.

Semi-partial correlations were used to examine the relationship between performance

on the English 12 examination (selection and supply items) and test-wiseness, adjusted for verbal ability. Given differences in the marking procedures for the short-answer and extended-answer subtests, these subtests were analyzed separately.

In addition, test-wiseness was examined with respect to differences of level and variability of performance on the English 12 examination (selection, short-answer, and extended-answer subtests) between test-wise ( $n=61$ ) and test-naive ( $n=76$ ) samples, adjusting for verbal ability. Group membership was determined by TTW test scores, with test-wise subjects scoring above 17 and test-naive subjects scoring below 11. Group means were analyzed using a one-way Analysis of Covariance (ANCOVA), while differences of variability of performance were analyzed using an F-test for independent variances.

Previous research suggested that the answers to the questions concerning the selection type items, but excluding the notion of variability, should be positive (Ardiff, 1965; Benson, 1988; Dreisbach & Keogh, 1982; Fueyo, 1977; Kuntz, 1981; Prell & Prell, 1985; Sarnacki, 1979). The answers to the questions concerning supply type items and those concerning the notion of variability were less certain due to the lack of research with items written in the supply format.

## **5.2 Summary and Discussion of the Results**

### **5.2.1 Substudy I**

The semi-partial correlations between the scores on the selection, short-answer, and extended-answer subtests of the English 12 examination and the residual on test-wiseness following the removal of verbal ability were positive. Two of these correlations were found to be statistically significant ( $p < 0.01$ ), thus approximately four percent of the variance was uniquely attributable to test-wiseness for the English 12 selection and short-answer

subtest. Although the correlations were small, this finding indicates that test-wiseness accounts for at least some of the difference between higher and lower scoring students, following the removal of verbal ability, as suggested by Bajtelsmit (1975), Diamond and Evans (1972), and Rowley (1974).

Analyses of the standardized residuals of test-wiseness on performance on the English 12 examination for the three subtests, adjusted for verbal ability, confirmed linearity. There were some instances of outliers ( $\pm 2.5$  standard deviations) which were not deleted as the sample and population characteristics were similar.

The results of the present study support previous research whereby test-wiseness was found to account for a unique portion of variance in test scores for a variety of tests (Callenbach, 1974; Millman & Setijadi, 1966; Oakland, 1972; Omvig, 1971; Rowley, 1974).

### 5.2.2 Substudy II

The results of the one-way ANCOVA revealed that the adjusted means of the test-wise subsample were statistically greater than the corresponding means for the test-naive subsample on the selection and short-answer subtest. There was, however, no difference between the means of these two groups on the extended-answer subtest.

Lack of significant difference between the test-wise and test-naive samples for the extended-answer subtest, may be attributable to the way in which this subtest was scored. In contrast to the exact scoring for selection and short-answer subtests, a holistic marking approach was used to score the extended-answer subtest. Further, the extended-answer subtest may have assessed a skill, or skills, (related specifically to writing or composition ability only) different from that of the selection and short-answer subtests.

Differences between variabilities of the test-wise and the test-naive samples for the

three subtests were not statistically different. This does not support Gibb's (1964) research whereby a significantly larger amount of variability for the test-wise group, as compared to the test-naive group, was found.

### **5.3 Limitations**

The generalizability of this study is limited to the method by which the extended-answer subtest was scored. Use of a holistic marking guide in contrast to a more objective marking system may yield results comparable to those found for the selection and short-answer, objectively scored items. Further, the TTW may not have assessed test-wiseness skills which were unique to the extended-answer subtest. The TTW was designed to assess attainment of four selection-type skills and, as such, may not have adequately addressed differences in test-wiseness between students for this subtest.

### **5.4 Further Discussion**

Results from Substudy I and Substudy II showed that test-wiseness does influence performance on the English 12 examination for the selection and short-answer items, when considered simultaneously with verbal ability. These results support previous research which used selection items on various standardized measures (Benson, 1988; Dreisbach & Keogh, 1982; Fueyo, 1977; Sarnacki, 1979; Wahlstrom & Boersma, 1968), although verbal ability was not accounted for prior to this study. However, test-wiseness does not influence performance on the extended-answer item of the English 12 examination. Differences in marking procedure may account, in part, for lack of differences in the findings.

The relationship between knowledge of test-wiseness skills and the ability to apply that knowledge to questions which require subject area knowledge may partially explain

the low internal consistency of the TTW ( $\text{rel}=0.37$ ). Students may understand the various test-wiseness strategies as outlined by Millman et al. (1965), but may be unable to apply these strategies when their subject area knowledge does not match the knowledge required. Many of the questions on the TTW may have required subject area knowledge (cognitive abilities) beyond that possessed by many of the subjects writing the TTW. Students who scored in the mid-range may have had knowledge of test-wiseness strategies but lacked the content knowledge necessary to successfully apply these strategies. Thus, knowledge of a certain test-wiseness skill may be independent of the ability to apply that skill. This notion is supported by the work of Rogers and Bateson (1990b).

Similarly, random responses to questions would contribute to the unreliability of the TTW. The model of test-taking proposed by Rogers and Bateson (1990a) contains four points where students could exit and present an answer. These four points correspond to four response strategies: a simple random guess, an “educated” random guess, a “test-wise” guess, or selection of the correct option. Clearly, two of these responses involve random guessing which contributes to measurement error and, correspondingly, the unreliability of the TTW. This would then attenuate the correlations between performance on the English 12 examination and the TTW. However, given the model, correction for attenuation is not warranted in this case (Rogers & Bateson, 1990a). Consequently, the low to moderate correlations observed are likely as high as one can get given the solution strategies used by the students.

## 5.5 Recommendations

### 5.5.1 Implications for Practice

Although low, the semi-partial correlations  $\rho_{E(T.V)_{SEL}}$  and  $\rho_{E(T.V)_{SA}}$  for the selection and short-answer subtests of the English 12 Provincial Examination indicate that test-wiseness is a factor which must be considered by the Ministry of Education in British Columbia, teachers of grade twelve students, and grade twelve students themselves. Further, results from Substudy II show test-wiseness skills were not uniformly possessed by all. Consequently, to the extent test-wiseness can be applied, those low in test-wiseness will be penalized in comparison to those high in test-wiseness.

The manner and extent to which test-wiseness operates on items on the provincial examinations needs to be addressed. Inasmuch as those involved in construction of the provincial examinations hope test items are measuring the students' knowledge of the subject matter involved on the test rather than test-wiseness, item susceptibility to test-wiseness and the possession of differential test-wiseness ability by test takers poses problems in test score interpretation.

Two suggestions, which, if taken, would eliminate the adverse effects of test-wiseness are:

- I. change the format of the examination by producing items which are not susceptible to test-wiseness; and
- II. provide test-wiseness materials to teachers and students as part of a test preparation package to be used when studying for the English 12 examination.

While producing items which are not susceptible to test-wiseness cues is difficult, (Benson, 1988; Sarnacki, 1979) Standard 3.11 of the *Standards for Educational and Psychological Testing* (American Psychological Association, 1985, p.27 - 28) is clear with



respect to providing test-wiseness instruction:

STANDARD 3.11 – When test-taking strategies that are unrelated to the constructs or content measured are found to influence test performance significantly, these strategies should be explained to test takers before the test is administered either in an information booklet or, if the explanations can be made briefly, along with the test directions. The use of such strategies by all test takers should be encouraged if their effect facilitates performance and discouraged if their effect interferes with performance. (*Primary*).

### 5.5.2 Further Research

Results from this study present several possibilities for further research with the expansion of the definition of test-wiseness and in its replication in other examinable subject areas.

It was noted that test-wiseness had a statistically significant impact upon the multiple-choice and short-answer subtests. Thus, it is suggested that further research on test-wiseness expand upon the definition and/or taxonomy of test-wiseness to incorporate short-answer item types. A possibility for further research is the development of a measure using short-answer test-wiseness susceptible items to directly assess this test-wiseness skill in students.

For this study test-wiseness was considered to be “logically independent of the examinee’s knowledge of the subject matter for which the items are supposedly measures” (Millman et al., 1965, p.707). Though this definition appears easy to understand, it is difficult to assess. How does one separate knowledge of test-wiseness strategies from subject matter (content) knowledge? Is it possible to separate the them? If we can accurately and adequately separate them, will we be able to produce a measure to assess

test-wiseness? Research into a different methods of measuring test-wiseness is warranted. Expanding the definition of test-wiseness to include the notion of a dichotomy, independent of content knowledge, is suggested.

Further, the use of internal consistency estimates to measure the reliability of test-wiseness instruments may not be appropriate. Stability in attainment of test-wiseness skills has been shown by Rogers and Bateson (1990b) in their validation study. Students who received scores of 18+ on the TTW received high scores on a similar type of measure of test-wiseness, with those scoring below 11 on the TTW receiving low scores on this instrument as well. Thus, our methods of assessing the reliability of measures of test-wiseness may be inadequate and inaccurate. Using parallel forms, or some other measure of stability, may be more accurate in our assessment of test-wiseness skills.

It is also suggested that the definition of test-wiseness include various factors of test-wiseness, such as verbal ability, possibly in the form of a linear combination. Test-wiseness may possibly be a specialized form of verbal ability. Research into the above possibilities, either in the form of a single study comparing the feasibility of the two definitions, or individually, is suggested.

As this study is the first to address the effect of test-wiseness on supply items but limits itself to the English 12 examination, the present study needs to be replicated in other subject areas. Results for the English 12 examination were significant; however, the same may, or may not, be true for other provincial examinations on subjects such as Biology, Geography, or History. Similarly, where government testing programs for grade 12 students exist in other provinces, these tests should be examined for test-wiseness cues and appropriate measures taken if such cues are present.

Although the extended-answer subtest was not affected by test-wiseness ability, this does not guarantee similar extended-answer items will be immune from such ability.

Inasmuch as other extended-answer items are marked using approaches other than the holistic approach used in the present study, replication of this study with essays scored in alternative ways is warranted.

## Bibliography

- [1] American Psychological Association. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- [2] Ardif, M. B. (1965). *The relation of three aspects of test-wiseness to intelligence and reading ability in grades three and six*. Unpublished master's thesis, Cornell University, Ithaca, NY.
- [3] Anastasi, A. (1981). Diverse effects of training on tests of academic intelligence. *New Directions for Testing and Measurement*, 11, 5-19.
- [4] Anderson, J. (1979). Intercorrelations of EPT scores, grade 12 English grades, and first year English grades for 1978/1979 students at two universities. *EPT Report 4:1* Vancouver: Educational Research Institute of British Columbia.
- [5] Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. C. (1983). Effects of coaching programs on achievement test performance. *Review of Educational Research*, 53(4), 571-585.
- [6] Bateson, D. (1984, June). *Provincial examinations in British Columbia*. Paper presented at the annual meeting of the Canadian Educational Research Association, Guelph, Ontario.
- [7] Benson, J. (1988). *The psychometric and cognitive aspects of test-wiseness: A review of the literature*. University of Maryland. Manuscript submitted for publication.
- [8] Benson, J., Urman, H., & Hocevar, D. (1986). Effects of test-wiseness training and ethnicity on achievement of third- and fifth-grade students. *Measurement and Evaluation in Counseling and Development*, 18(4), 154-162.
- [9] Callenbach, C. (1973). The effects of instruction and practice in content-independent test-taking techniques upon the standardized reading scores of selected second-grade students. *Journal of Educational Measurement*, 10(1), 25-30.
- [10] Crehan, K. D., Gross, L. J., & Slakter, M. J. (1978). Developmental Aspects of test-wiseness. *Educational Research Quarterly*, 3(1), 40-44.
- [11] Crehan, K. D., Kohler, R. A., & Slakter, M. J. (1974). Longitudinal studies of test-wiseness. *Journal of Educational Measurement*, 11(2), 209-212.
- [12] Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.

- [13] Diamond, J. J. & Evans, W. J. (1972). An investigation of the cognitive correlates of test-wiseness. *Journal of Educational Measurement*, 9(2), 145-150.
- [14] Dolly, J. P., & Williams, K. S. (1983, October). *Teaching testwiseness*. Paper presented at the annual meeting of the Rocky Mountain Educational Research Association, Jackson, Wyoming.
- [15] Dreisbach, M. & Keogh, B. K. (1982). Testwiseness as a factor in readiness test performance of young Mexican-American children. *Journal of Educational Psychology*, 74(2), 224-229.
- [16] Educational Measurement Research Group (1988). *Language Proficiency Index: 1989 LPI registration form*. Vancouver, British Columbia: Author.
- [17] Educational Measurement Research Group (1988). *Language Proficiency Index (LPI): An overview form*. Vancouver, British Columbia: Author.
- [18] Educational Research Institute of British Columbia (ERIBC) (1980) Factorial validity of the English Placement Test. *ERIBC Report*, 80:7 Vancouver, British Columbia: Author.
- [19] Evans, F. R. & Pike, L. W. (1982). The effects of instruction for three mathematics item formats. *Journal of Educational Measurement*, 10(4), 257-271.
- [20] Fraser, P. (1984). News Release. *ERIBC Report*, Vancouver, British Columbia: Educational Research Institute of British Columbia.
- [21] Fraser, P. & Anderson, J. (1984). Information Bulletin. *ERIBC Report*, Vancouver, British Columbia: Educational Research Institute of British Columbia.
- [22] Frankel, E. (1960). Effects of growth, practice, and coaching on Scholastic Aptitude Test scores. *Personnel and Guidance Journal*, 38, 713-719.
- [23] French, J. W. & Dear, R. E. (1959). Effect of coaching on an aptitude test. *Educational and Psychological Measurement*, 29, 319-330.
- [24] Fueyo, V. (1977). Training test-taking skills: A critical analysis. *Psychology In The Schools*, 14(2), 180-185.
- [25] George, P. (1985). Coaching for tests: A critical look at the issues. *Curriculum Review*, 25(1), 23-27.
- [26] Gibb, B. G. (1964). *Test-wiseness as secondary cue response*. Doctoral Dissertation, Stanford University. (University Microfilms, 1964. No.64-7643)
- [27] Glass, G. V. & Hopkins, K. D. (1984). *Statistical Methods in education and psychology*. (2nd ed.) New York: Prentice Hall.
- [28] Kalechstein, M., Kalechstein, M., & Docter, R. (1981). The effects of instruction on test-taking skills in second-grade black children. *Measurement and Evaluation in Guidance*, 13(4), 198-202.

- [29] Kerlinger, F. N. & Pedhauser, E. J. (1973). *Multiple regression in behavioural research*. New York: Holt, Rinehart, & Winston.
- [30] Kuntz, P. A. (1981). *Test-wiseness cues in the options of mathematics items: Prevalence and trainability*. Unpublished master's thesis, Cornell University, Ithica, NY.
- [31] Learning Assessment Branch (1973). *The move away from examinations: What happens now?* (Memorandum for the Ministry of Education). Victoria, B.C.
- [32] Millman, J. (1966). *Test-wiseness in taking objective achievement and aptitude examinations*. Final Report. College Entrance Examination Board.
- [33] Millman, J., Bishop, C. H., & Ebel, R. (1965). An analysis of test-wiseness. *Educational and Psychological Measurement*, 25(3), 707-726.
- [34] Millman, J. & Setijadi (1966). A comparison of the performance of American and Indonesian students on three types of test items. *Journal of Educational Research*, 59(6), 273-275.
- [35] Ministry of Education. (1983). Ministry Policy Circular. Victoria, B. C.: Author.
- [36] Ministry of Education: Student Assesment Branch: Province of British Columbia (1989). *English 12 provincial examinations essay: Holistic scoring procedures*. Queen's Printers #79027. Victoria: Author.
- [37] Moore, J. C., Schutz, R. E., & Baker, R. L. (1966). The application of self-instructional technique to develop a test-taking strategy. *American Educational Research Journal*, 3(2), 13-17.
- [38] Oakland, T. (1972). The effects of test-wiseness material on standardized test performance of preschool disadvantaged children. *Journal of School Psychology*, 10(4), 355-360.
- [39] Omvig, C. P. (1971). Effects of guidance on the results of standardized achievement testing. *Measurement and Evaluation in Guidance*, 4(1), 47-52.
- [40] Pallone, N. J. (1961). Effects of short-term and long-term developmental reading courses upon the SAT verbal scores. *Personnel and Guidance Journal*, 39, 654-657.
- [41] Pedhauser, E. J. (1982). *Multiple regression in behavioural research*. (2nd ed.) New York: Holt, Rinehart, & Winston.
- [42] Petty, N. E. & Harrell, E. H. (1977). Effects of programmed instruction related to motivation, anxiety, and test wiseness on group IQ test performance. *Journal of Educational Psychology*, 69(5), 630-635.
- [43] Prell, J. M. & Prell P. A. (1986, November). *Improving test scores - Teaching test-wiseness: A review of the literature*. Research Bulletin.

- [44] Powers, D. E. & Alderman, D. L. (1983). Effects of test familiarization on SAT performance. *Journal of Educational Measurement*, 20(1), 71-79.
- [45] Putnam, J. H. & Weir, G. M. (1925). *Survey of the school system*. Victoria, B. C.: King's Printer.
- [46] Rogers, W. T. & Bateson, D. J. (1990a). *The impact of test-wiseness upon year-end school leaving examinations*. Manuscript submitted for publication.
- [47] Rogers, W. T. & Bateson, D. J. (1990b). *Verification of a model of test-taking behaviour of high school seniors' examinations*. Manuscript submitted for publication.
- [48] Rowley, G. L. (1974). Which examinees are most favoured by the use of multiple choice tests? *Journal of Educational Measurement*, 11(1), 15-23.
- [49] Samson, G. E. (1985). Effects of training in test taking skills on achievement test performance: A quantitative synthesis. *Journal of Educational Research*, 78(5), 261-266.
- [50] Sarnacki, R. E. (1979). An examination of test-wiseness in the cognitive test domain. *Review of Educational Research*, 49(2), 252-279.
- [51] Sax, G. (1974). *Principles of educational measurement and evaluation*. Belmont, CA: Wadsworth Publishing Company, Inc..
- [52] Slakter, M. J., Kohler, R. A., & Hampton, S. H. (1970a). Grade level, sex, and selected aspects of test-wiseness. *Journal of Educational Measurement*, 7(2), 119-122.
- [53] Slakter, M. J., Kohler, R. A., & Hampton, S. H. (1970b). Learning test-wiseness by programmed texts. *Journal of Educational Measurement*, 7(4), 247-254.
- [54] Smith, J. K. (1982). Converging on correct answers: A peculiarity of multiple choice items. *Journal of Educational Measurement*, 19(3), 211-220.
- [55] Stanley & Hopkins. (date unknown). Factors that influence test performance. *Educational Measurement*, a reprinted draft copy.
- [56] Stanley, J. C. (1971). Reliability in R. L. Thorndike (ed.) *Educational measurement*. (2nd ed.) Washington, DC: American Council on Education.
- [57] Swinton, S. S. & Powers, D. E. (1983). A study of the effects of special preparation on GRE analytical scores and item types. *Journal of Educational Psychology*, 75(1), 104-115.
- [58] Thorndike, R. L. & Hagen, E. (1969). *Measurement and evaluation in psychology and education* (p.95). New York: John Wiley & Sons Inc..

- [59] Wahlstrom, M. & Boresma, F. J. (1968). The influence of test-wiseness upon achievement. *Educational and Psychological Measurement*, 28, 413-420.
- [60] Williams, K. S. & Dolly, J. P. (1983, April). *Can teaching cognitive strategies improve test-taking performance?* Paper presented at the annual meeting of the Rocky Mountain Psychological Association, Snowbird, Utah.



## Appendix A

The English Twelve Provincial Examination

### PROVINCIAL EXAMINATION

• MINISTRY OF EDUCATION •

# ENGLISH 12

#### GENERAL INSTRUCTIONS

1. Insert the sticker with your Student I.D. Number in the allotted space above.  
**Under no circumstances is your name or identification, other than your Student I.D. Number, to appear on this paper.**
2. Take the separate Answer Sheet and follow the directions on its front page.
3. Be sure you have a READINGS BOOKLET which contains the prose and poetry passages you will need to answer the questions in this booklet.
4. Be sure that you have an HB pencil and an eraser for completing your answer sheet. Follow the directions on the answer sheet when answering multiple-choice questions.
5. 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**.
6. 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.

© 1989 Ministry of Education

## PART A: EDITING SKILLS

Value: 10 marks (one mark per question)

Suggested Time: 15 minutes

**INSTRUCTIONS:** Read the passage located on the left side of pages 1 and 2. For questions 1 to 10, choose the word, phrase, or punctuation which **BEST** completes each blank in the passage and record your choices on the answer sheet provided. Using an HB pencil, completely fill in the circle that has the letter corresponding to your answer.

## MONKEY BUSINESS

The darker recesses of the Vancouver Aquarium gave way to the bright sunlight of a glorious (1). Nevertheless, I felt (2), and I knew it was for the following reason (3) it saddened me to see wild creatures removed from their natural habitat to perform unnatural acts in captivity. Should we enjoy the spectacle of majestic killer whales leaping fifteen feet out of the water to ring a bell on the end of a pole? Oh, yes, I know their willingness and their enjoyment (4) that no cruelty is used in their training. However, such artificial stunts, no matter if performed by a whale, or a smaller creature such as a porpoise, (5) demeaning, robbing them of their natural dignity. (6), for some people, such performances are immensely satisfying.

1. A. spring day in april  
B. spring day in April  
C. Spring day in April  
D. Spring Day in April
2. A. distraught  
B. disdainful  
C. distracted  
D. disinterested
3. A. : (colon)  
B. , (comma)  
C. ; (semi-colon)  
D. no punctuation
4. A. indicate  
B. indicates  
C. is indicative  
D. seems to indicate
5. A. is  
B. are  
C. was  
D. were
6. A. Yet  
B. Thus  
C. Moreover  
D. Furthermore

The \_\_\_\_ (7) \_\_\_\_ for whom such activities are a source of pleasure may be found outside of the monkey cage in Stanley Park zoo pulling faces, jumping up and down, and \_\_\_\_ (8) \_\_\_\_ scratching motions under his arms. Suddenly, I was shaken out of my musings by one of this very type. "See that orangutan lying on that \_\_\_\_ (9) \_\_\_\_ one of these persons to me. "It reminds me of my ol' man, the way he's stretched out jus' like in front of the TV." When I looked at the orangutan and back to the speaker, I thought I could see a genetic \_\_\_\_ (10) \_\_\_\_.

7. A. type of person  
B. type of persons  
C. types of person  
D. types of persons
8. A. making  
B. is making  
C. are making  
D. may be making
9. A. tree" said  
B. tree", said  
C. tree?" said  
D. tree"? said
10. A. flaw  
B. residue  
C. contrast  
D. connection

**PART B: READING COMPREHENSION**

Value: 18 marks

Suggested Time: 35 minutes

---

**INSTRUCTIONS:** Read "Increasing Life Expectancy Means Growing Role For Euthanasia" on pages 1 to 3 of the READINGS BOOKLET. Section One questions must be answered on the answer sheet, and Section Two questions must be answered in the spaces provided on pages 5 and 6. Follow instructions carefully.

---

## Section One (8 marks - one mark per question)

---

**INSTRUCTIONS:** Select the **BEST** answer for each question and record it on the answer sheet provided.

---

11. In paragraph four, "categorical" means
- A. absolute.
  - B. essential.
  - C. ambiguous.
  - D. discriminatory.
12. As used in paragraph seven, "purgatory" means
- A. a state of confusion.
  - B. the welcome relief of pain.
  - C. a state of continual suffering.
  - D. the welcome release of death.
13. Turning off life-supporting machines so that a terminally ill patient could die would be considered a form of
- A. suicide.
  - B. palliative care.
  - C. active euthanasia.
  - D. passive euthanasia.
14. Palliative treatment is not widespread at present because
- A. we are not ready to accept euthanasia.
  - B. the elderly do not find it an acceptable alternative.
  - C. society has a preference for other methods of treatment.
  - D. the facilities required are expensive and not readily available.
15. The degree of support for euthanasia outside of France is
- A. not documented.
  - B. as strong as it is in France.
  - C. blocked by religious organizations.
  - D. not as strong as it is within France.

16. Palliative treatment, suggests the author, might be
  - A. ineffectual in treating incurable diseases.
  - B. a way of preventing euthanasia altogether.
  - C. merely a subtle, slower form of euthanasia.
  - D. too stressful for patients and their families.
  
17. The author implies that euthanasia
  - A. should be made legal.
  - B. should be prevented by law.
  - C. is a difficult and complicated issue.
  - D. is a complex and immoral procedure.
  
18. Throughout this article the author maintains an attitude towards his subject which can be **BEST** described as
  - A. calm and rational.
  - B. intense and emotional.
  - C. diplomatic and unprincipled.
  - D. complacent and uncommitted.

**WRITE IN INK****INCREASING LIFE EXPECTANCY MEANS GROWING ROLE FOR EUTHANASIA**  
(Pages 1 to 3 of the READINGS BOOKLET)**Section Two (10 marks)**

1. According to the author, what are **TWO** reasons for the lack of a common standard or approach to euthanasia? **(2 marks)**

(a) \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

(b) \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

Question 1. \_\_\_\_\_ 1.  
(2)

2. Identify **TWO** reasons which the author suggests will increase the demand for euthanasia in the future. **(2 marks)**

(a) \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

(b) \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

Question 2. \_\_\_\_\_ 2.  
(2)

3. State **THREE** arguments against legalizing euthanasia which appear in this article. (3 marks)

(a) \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

(b) \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

(c) \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

-----  
Question 3. Score: \_\_\_\_\_ 3.  
(3)  
-----

4. The author of this article frequently uses examples to clarify or to support his discussion. List **THREE** other DEVICES or TECHNIQUES which he uses to present ideas or arguments. (3 marks)

(a) \_\_\_\_\_

\_\_\_\_\_

(b) \_\_\_\_\_

\_\_\_\_\_

(c) \_\_\_\_\_

\_\_\_\_\_

-----  
Question 4. Score: \_\_\_\_\_ 4.  
(3)  
-----

**PART C: POETRY**

Value: 19 marks

Suggested Time: 30 minutes

**INSTRUCTIONS:** Read "A Kite is a Victim" on page 4 of the READINGS BOOKLET. Section One questions must be answered on the answer sheet, and Section Two questions in the spaces provided on pages 8 and 9. Follow instructions carefully.

---

**Section One (5 marks - one mark per question)**

---

**INSTRUCTIONS:** Select the **BEST** answer for each question and record it on the answer sheet provided.

- 
19. Lines 13 and 14 contain examples of
- A. assonance and simile.
  - B. metaphor and hyperbole.
  - C. onomatopoeia and alliteration.
  - D. alliteration and personification.
20. Which figure of speech is used in the following quotation: "then you pray the whole cold night before,/under the travelling cordless moon"?
- A. Irony
  - B. Simile
  - C. Metaphor
  - D. Apostrophe
21. Which quotation **BEST** illustrates a combination of controlled and free flight?
- A. "it lives like a desperate trained falcon/in the high sweet air" (lines 6 and 7)
  - B. "and you can always haul it down/to tame it in your drawer" (lines 8 and 9)
  - C. "a kite is the last poem you've written/so you give it to the wind" (lines 15 and 16)
  - D. "you don't let it go/until someone finds you/something else to do" (lines 17 to 19)
22. In all four stanzas of this poem, the poet suggests a contrast between
- A. virtue and rejection.
  - B. spirituality and turmoil.
  - C. spontaneity and restraint.
  - D. innocence and worldliness.
23. The mood created by this poem is one of
- A. joy and exhilaration.
  - B. domination and oppression.
  - C. whimsy and sentimentality.
  - D. determination and frustration.



## WRITE IN INK

---

"A Kite is a Victim" (Page 4 of the Readings Booklet)

---

## Section Two (14 marks)

- 
5. Explain why the use of free verse is appropriate to the subject of the poem. (2 marks) Answer in complete sentences.

---

---

---

---

---

---

---

---

Questions 5. Score: \_\_\_\_\_ 5.  
(2)

---

6. Explain how it is possible for a kite to pull "gentle enough to call you master, strong enough to call you fool." (2 marks) Answer in complete sentences.

---

---

---

---

---

---

---

---

Question 6. Score: \_\_\_\_\_ 6.  
(2)

---

---

Complete Sentences: \_\_\_\_\_ 7.  
(1)

---

7. In line one, Cohen says that "a kite is a victim." In paragraph form, quote and fully explain **THREE** other metaphors involving a kite. Refer to the **LAST THREE STANZAS** only. **(9 marks)** Three marks will be awarded for the quality of your written expression.

[illegible]

Question 7. Score:  $\frac{\quad}{(6)}$  8.

Written Expression:  $\frac{(9)}{(3)}$  9.

**PART D: PROSE****Value: 29 marks****Suggested Time: 45 minutes**

**INSTRUCTIONS:** Read the story entitled "The Kool-Aid Wino" on pages 5 and 6 of the READINGS BOOKLET. Section One questions must be answered on the answer sheet, and Section Two questions in the spaces provided on pages 12 to 14. Follow instructions carefully.

---

**Section One (5 marks - one mark per question)**

---

**INSTRUCTIONS:** Select the **BEST** answer for each question and record it on the answer sheet provided.

- 
24. In the quotation "whose wet diapers were in various stages of anarchy," the word "anarchy" means
- A. discord.
  - B. disorder.
  - C. disunion.
  - D. disgrace.
25. The Kool-Aid wino's family may be **BEST** described as
- A. poor, but proper.
  - B. poor and tattered.
  - C. poor and close-knit.
  - D. poor, but well-housed.
26. The character of the grocer is **BEST** described as
- A. flat.
  - B. round.
  - C. motivated.
  - D. developing.
27. The simile, "like a famous brain surgeon removing a disordered portion of the imagination" means that the Kool-Aid wino
- A. is famous for his Kool-Aid making skills.
  - B. makes Kool-Aid with skill and precision.
  - C. enters another reality when making Kool-Aid.
  - D. is a person who prefers to exist in an orderly fashion.
28. The setting of this story is
- A. on a large farm.
  - B. in an urban centre.
  - C. on a remote ranch.
  - D. in a rural community.

**WRITE IN INK****"The Kool-Aid Wino" (Pages 5 and 6 of the READINGS BOOKLET)****Section Two (24 marks)**

8. It is clear that the narrator is interested in the Kool-Aid wino and his eccentric behaviour. Identify and explain **THREE** other attitudes of the narrator toward the Kool-Aid wino implied in the story. Each explanation must include a specific reference to the story. **(6 marks)**

(a) \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

(b) \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

(c) \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

-----  
Question 8. Score: \_\_\_\_\_ 10.  
(6)  
-----

- 
- This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

Question 9. Score:  $\frac{\quad}{(6)}$  11.

Written Expression:  $\frac{\quad}{(3)}$  12.

10. In the story's title, Brautigan refers to the main character as a "wino." In paragraph form, describe **THREE** examples of behaviour which support his being called a wino. (9 marks)  
Three marks will be awarded for the quality of your written expression.

This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

Question 10. Score:  $\frac{\quad}{(6)}$  13.

Written Expression:  $\frac{(3)}{(3)}$  14.

**PART E: COMPOSITION**

**Value: 24 marks**

**Suggested Time: 55 minutes**

---

**INSTRUCTIONS:** Using standard English, write a coherent, unified, multi-paragraph composition of 300-500 words on **ONE** of the topics listed below. In your composition you may use the techniques of exposition, narration, or description which you feel are appropriate to your topic. Do not write on more than one topic; if you do so, only the first will be marked.

In planning your composition, make sure you use skills of organization and development appropriate to the type of composition you choose.

Use the pages headed "**Organization and Planning**" for your rough work. Write your composition **IN INK** on the pages headed "**Finished Work**."

---

**TOPICS**

11. Write a multi-paragraph composition using **ONE** of the following topics:
- A. Friends and relatives.
  - B. Fans.
  - C. Music to my ears.

## Appendix B

### English 12 Composition – Scoring Guide

A composition may or may not show all of the features of any one scale point.

5 – 6

UPPER THIRD PAPERS have original and well integrated ideas, are clear and controlled with a mature sense of voice and awareness of audience. Organization is natural and unobtrusive. Sentences are varied, diction effective and errors insignificant

#### EXCELLENT

The EXCELLENT is original, refreshing, vigorous, or interesting.

An eccentric paper may signal artistry. Supporting detail is  
6 mature and informed. A sense of voice and audience awareness immediately engages the reader. Expression flows effortlessly into a literate, integrated whole. As a first draft, it is virtually error-free.

#### PROFICIENT

The PROFICIENT paper is highly readable, but conventional.

There is some originality. The paper is engaging but not as



- 5 consistently engaging as a 6. Supporting details are appropriate. Organization is consistent in direction and tone. sentences are controlled, vocabulary varied and errors unobtrusive.

3 – 4

MID-LEVEL PAPERS have conventional but uninspired ideas, and range from satisfactory to merely acceptable in coherence and development. Organization lacks focus and expression is awkward, repetitive and noticeably flawed.

### SATISFACTORY

- The SATISFACTORY paper is adequate, and workmanlike, but the ideas are not memorable. Development and supporting details are
- 4 competently but not expertly handled. Expression is appropriate, controlled and conventional, but lacks sophistication and colour. Errors neither overwhelm the reader nor distort the writer's purpose.

### LIMITED

- The limited paper is barely acceptable. Content is ill-defined, dull, uninspired and juvenile. It lacks focus;
- 3 examples may be trite, simplistic or random. Paragraphs, when present, lack structure; transitions are weak to non-existent.

Sentences lack variety, diction is repetitive, and errors obscure meaning.

LOWER THIRD PAPERS reveal a writer who has difficulty responding to the topic at any level. Examples are thin; development is inconsistent or faulty; expression is flawed to the point of distraction or lack of comprehension. Paper may be short and undeveloped.

### UNSATISFACTORY

The UNSATISFACTORY paper is underdeveloped, incomplete or superficial. Point of view and sense of audience are unclear or  
2 non-existent. Examples and details are inappropriate. Expression is awkward and depends on colloquialisms. English idiom is uncontrolled, diction inadequate and errors frequent.

### UNACCEPTABLE

The UNACCEPTABLE has neither purpose or focus. It may be too brief to allow development of ideas. Unity, coherence and  
1 emphasis are virtually non-existent. Organization, where present, is illogical, confusing or uncontrolled. Errors in standard English make ideas difficult to understand.

CANNOT BE EVALUATED

0 \*This is a special category reserved for papers which CANNOT BE EVALUATED. Some text has been produced, but the effort is characterized by one or more of the following:

- a) no discernible grasp of English idiom;
- b) too deficient in length to evaluate;
- c) errors make the paper unintelligible;
- d) the paper addresses a topic not given;

\*A zero can only be assigned by the marking chair or designate.

Papers which are left blank or contain only one or two words or a brief, incoherent phrase are given a mark of NR for NO RESPONSE.

Taken from: Ministry of Education: Student Assessment (1988). *English 12 provincial examinations essay: Holistic scoring procedures*. p.4.  
Victoria: Queen's Printer.

## Appendix C

### Rules for Holistic Scoring

a. Read quickly and score

Speed, reliability and accuracy of scoring will be enhanced if the following rules are followed:

\*avoid rereading

\*record score immediately

b. Do not use a marking pen

\*marking individual errors is too slow

\*marking individual errors tends to over-emphasize specific traits

c. Do not second-guess a score

\*first impressions tend to be more reliable

d. If unsure

\*refer to the descriptive guide

\*check anchor papers

\*give the paper to another marker

e. Remember: any one bundle could contain all top papers or

all bottom papers; therefore avoid creating a "normal" distribution of scores.

f. Remind yourself of biasing factors.

g. Remind yourself that this writing is the result of a

55 - minute test under exam conditions. In other words, you are marking a **FIRST DRAFT** with some quick revisions and corrections.

h. Initially, try classifying papers as a two-step process:

- \*Classify a paper either as upper half of the guide or lower half.
- \*Place the paper within scale point in the appropriate half.

Taken from: Ministry of Education:Student Assessment (1988). *English 12 provincial examinations essay: Holistic scoring procedures.* p.26-27.  
Victoria: Queen's Printer.

## Appendix D

### Test of Test-Wiseness

#### Background Section:

You should have received a computer readable answer sheet along with this test booklet. When answering on the answer sheet, **use a dark HB pencil only and darkly fill in the appropriate round bubble.** If you wish to change an answer, **erase** all traces of the wrong mark, then darken the correct response.

1. Print your **NAME** and the name of your **School** at the top of this Test Booklet.
2. Print your **NAME** (**Last Name** followed by a **BLANK** followed by **First Name**) in the name field boxes of the answer sheet and then fill in the appropriate bubble under each letter.
3. Fill in the appropriate circle in the field marked **SEX** (M or F).
4. Fill in your **BIRTH DATE**.
5. Please answer the following 3 questions (K, L and M) under the field marked **SPECIAL CODES**.

*In the **SPECIAL CODES 'K'** column fill in the appropriate bubble for the following question.*

**K. To which ethnic or cultural group do you belong?**

- |   |               |
|---|---------------|
| 0 | English       |
| 1 | French        |
| 2 | Native Indian |
| 3 | East Indian   |
| 4 | Chinese       |
| 5 | German        |
| 6 | Italian       |
| 7 | Japanese      |
| 8 | Vietnamese    |
| 9 | Other         |

*In the **SPECIAL CODES 'L'** column fill in the appropriate bubble for the following question.*

**L. Have you ever had any coaching or specific lessons on how to take a test?**

- |   |                          |
|---|--------------------------|
| 0 | No, never                |
| 1 | Yes, once or twice       |
| 2 | Yes, three or more times |

*In the **SPECIAL CODES 'M'** column fill in the appropriate bubble for the following question.*

**M. Have you ever practised writing provincial examinations using previous provincial examinations or questions from these examinations?**

- |   |                          |
|---|--------------------------|
| 0 | No, never                |
| 1 | Yes, once or twice       |
| 2 | Yes, three or more times |

## TEST OF TEST-WISENESS

This is a test of test-wiseness which measures some of the abilities needed to do well on tests. Many of the questions are about things you may not have studied. However, there are test-taking strategies which can be used to figure out what to do when faced with such questions.

For example:

**The greatest advantage of using slent in the manufacture of steel is that slent makes steel**

- a. transparent.
- b. stainless.
- c. heavy.
- d. rubbery.

Using test-wiseness strategies, options 'a' and 'd' can be eliminated since they are clearly not correct (steel is not transparent, nor is it rubbery). Therefore, either 'b' or 'c' is the correct answer. Now we stand a better chance of guessing the correct answer for we have narrowed the number of possible options down to two from four.

*Please be sure to follow the specific instructions for each of the following sections.*

### Section 1: Suggested Time: 20 minutes

**INSTRUCTIONS:** For each question, select the **BEST** answer and record your choice on the answer sheet provided. Each question is worth one mark. There will be no correction for guessing.

1. Compared to normal cells, bileuvial cells
  - a. divide more rapidly.
  - b. divide more slowly.
  - c. have more cytoplasm.
  - d. have more mitochondria.
2. The Flying Spider is known for its ability to
  - a. blend in with its surroundings.
  - b. glide through the air.
  - c. kill its prey with poison.
  - d. make very large webs.
3. The square root of 1.1 can be best approximated by
  - a. the cube root of 11056.
  - b. the solution of  $x^2 - 16 = 0$ .
  - c. using the binomial series.
  - d. factoring the expression  $9x^2 - 18$ .
4. Mr. Adams, in Henry Fiedling's Joseph Andrews,
  - a. learns his parents were of the nobility.
  - b. takes sick after fishing through the ice.
  - c. falls into the mud while reading.
  - d. discovers he is of noble birth.

5. In 404 B.C., the Athenian oligarchs, supported by Lysander and Theramenes,
  - a. approved a plan to rebuild sections of the city.
  - b. encouraged a military expedition against the Spartans.
  - c. rejected the constitution created by the committee of ten.
  - d. set up a commission of thirty to write a new constitution.
6. Which of the following would help to determine if D is the fourth harmonic of C with respect to A and B?
  - a. The relative size of angle ACB to angle ADB
  - b. The length of line segment AB
  - c. The fact that A, B, C, and D lie on one straight line
  - d. The straight line distance from A to B
7. A normal percentage of polymorphonuclear leukocytes found in the human peripheral blood is
  - a. 53/260.
  - b. 70%.
  - c. 115%.
  - d. 035.
8. The mercantilists believed in
  - a. freedom from all governmental interference in trade.
  - b. merchant control of colonies.
  - c. an income tax.
  - d. organized trade unions.
9. Wordsworth's "The Prelude" (1805)
  - a. tells of a descent into Hell in a Model-T Ford.
  - b. makes use of a distinction between "the sublime" and "the beautiful".
  - c. is concerned with the emerging African nations.
  - d. was influenced by Hemmingway's The Sun Also Rises.
10. The literature of the early eighteenth century is
  - a. public in nature, relating to society's outlooks and values.
  - b. private in nature, relating to an individual's emotions and feelings.
  - c. rough and irregular compared to the literature of the later eighteenth century.
  - d. filled with despair over the apparent collapse of traditional values.
11. Charles Dickens' Hard Times deals with
  - a. the difficult life of a factory worker.
  - b. the politics of the French chateau country.
  - c. the court of King Edward III.
  - d. the limitations of European existentialism.
12. Organisms of the Pavo genus
  - a. change from masculine to feminine gender.
  - b. display their plumage for the female.
  - c. become female after existing for a period of time as male.
  - d. possess an excess quantity of masculine hormone.
13. Which of the following most likely caused the War of 1693?
  - a. Spain was building roads to connect her cities.
  - b. France was going through great agricultural change.
  - c. France believed that Spain was increasing her troops.
  - d. Spain had a series of earthquakes.



14. A spherical triangle is the triangle on the surface of a sphere. What name is given to the number of degrees in a spherical triangle minus 180?
- The arc of the triangle.
  - The size of the triangle.
  - The spherical excess of the triangle.
  - The polar measurement of the triangle.
15. In Horace Walpole's The Castle of Otranto
- Manfred is the father of Hippolita.
  - Hippolita is the wife of Manfred.
  - Manfred is the uncle of Hippolita.
  - Hippolita is the daughter of Manfred.
16. The ring  $F[x] / s(x)$  is a field if and only if
- $s(x)$  is a prime polynomial over  $F$ .
  - $s(x)$  is a rational polynomial over  $F$ .
  - $F[x]$  is a multiple of  $s(x)$ .
  - any element of  $F[x]$  contains an inverse.
17. The career of Marius (155-86 B.C.), the opponent of Sulla, is significant in Roman history because
- he gave many outstanding dinners and entertainments for royalty.
  - he succeeded in arming the gladiators.
  - he showed that the civil authority could be thrust aside by the military.
  - he made it possible for the popular party to conduct party rallies outside the city of Rome.
18. A substance that, in its pure form, is the best conductor of electricity is
- water.
  - deuterium.
  - $H_2O$ .
  - silver.
19. The august character of the work of Pericles in Athens frequently causes his work to be likened to that in Rome of
- Augustus.
  - Sulla.
  - Pompey.
  - Claudius.
20. "Lucifer in Starlight" is
- a modern psychological story of World War II.
  - a Shakespearean sonnet by Gerard Manley Hopkins.
  - a controversial French novel written by Resnais.
  - the title of a Petrarchan sonnet by George Meredith.
21. The treaty of Brest Litovsk was ratified by Moscow because
- Tsar Alexander I wanted to prevent Napoleon's invasion of Russia.
  - Russia was unable to keep up with the armament manufacture of Austria.
  - Russia could not keep pace with the military production of Austria.
  - Nicolai Lenin wanted to get the Soviet Union out of World War I.

22. How many iambic feet (one iambic foot - one unstressed syllable followed by one stressed syllable, as in "perFORM") are in each line of Robert Pack's poem "The Compact"?
- a. 1
  - b. 5
  - c. 16
  - d. 22
23. What is the probability that a needle of length  $L < D$ , when dropped on a table ruled with equidistant parallel lines at distance  $D$  apart, will cross one of the lines?
- a.  $\frac{\sqrt{3}}{3} (LD)$
  - b.  $\frac{2L}{\pi D}$
  - c.  $L + .001D$
  - d.  $\frac{2D}{\pi L}$
24. The Feulgen Nuclear Reaction demonstrates the presence of
- a. desoxyribonucleoprotein.
  - b. lysosomes.
  - c. mitochondria.
  - d. endoplasmic reticulum.

---

Please Turn Page

---

## Appendix E

### Instrument Used For Validation of Test-Wiseness

(Rogers & Bateson, 1990b)

1. The compromise between the Democrats and Republicans after the post-civil war election of 1876 resulted in
  - (a) federal aid to Southern railroads.
  - (b) the entrance of Texas into the Union as a slave state.
  - (c) a treaty with Joseph Stalin.
  - (d) the Fugitive Slave Law.
2. If rubulose biphosphate were removed from a chloroplast, which of the following statements would BEST describe the immediate result?
  - (a)  $CO^2$  could not enter the Calvin cycle.

- (b) ATP could not be produced in the thylakoid.
  - (c)  $O^2$  could not enter the Calvin cycle.
  - (d) Light energy could not be trapped in the grana.
3. In 1911 the Tories in control of the House of Commons showed their political influence and democratic spirit when they
- (a) enacted the Asquith price-control bill.
  - (b) vetoed a bill to extend the suffrage privilege.
  - (c) forced the House of Lords to accept a reduction in its power.
  - (d) were defeated in an attempt to enact the National Insurance Act.
4. The sigma effect of the Fahraeus-Lindquist phenomenon is related to
- (a) the flow of liquid through the kidney.
  - (b) the muscular contractions of the kidney.

(c) the kidney's transmission of fluids.

(d) the diameter of the red cells flowing through the kidney's blood vessels.

5. Which of the following would cause an oat seedling to bend to the right?

(a) Auxins placed on the left side of the shoot tip.

(b) Giberellins place on the left side of the shoot tip.

(c) Cytokinins place on the right side of the shoot tip.

(d) Auxins placed on the right side of the shoot tip.

6. Why is Cavalieri's Principle important in Solid Geometry?

(a) It shows that the surface area of a square of side  $s$  is  $s^2$ .

(b) It provides contradictions to the principles of Euclid and Gauss.

- (c) It is used to prove that two polygons are congruent.
  - (d) It provides the basis for finding the volume formulae for many solids.
7. A major characteristic of natural resources is that they are
- (a) sources of energy.
  - (b) non-recyclable.
  - (c) non-renewable.
  - (d) unevenly distributed.
8. The emperor of the ancient Hsin Dynasty who resigned to undertake radical reform was
- (a) Saigon.
  - (b) Wang Mang.
  - (c) Mao T'se Tung.

(d) Alexander I.

9. The nose

(a) records light sensation.

(b) develops during gastrulation.

(c) has two moveable joints.

(d) is structured in part by the turbinals.

(e) is an organ of balance.

10. Stridulation at times facilitates

(a) coordination.

(b) dispersal.

(c) nutrition.

(d) excretion.

11. After the treaty of Rastatt and Baden in March – September 1714, Austria took possession of the Spanish Netherlands.

(a) True.

(b) False.

12. The Zmstvo-Law proclaimed by Alexander II of Russia in 1864

(a) freed millions of Russian serfs.

(b) abolished the old system of class courts.

(c) allowed justices of the peace to deal with minor civil suits.

(d) introduced the principle of universal military liability.

13. Every bounded infinite set has at least one cluster point.

(a) True.

(b) False.



14. "If the group  $G$  has order  $n$ , the order of every subgroup  $H$  of  $G$  is a divisor of  $n$ . " This is the theorem of

(a) Napoleon.

(b) Bach.

(c) Cauchy.

(d) Homer.

(e) Legrange.

## Appendix F

### Language Proficiency Index

#### Section 4: Sentence Structure

Some, though not all, of the ten items found below contain an error related to sentence structure, such as a misplaced modifier, a run-on sentence, a lack of subject-verb agreement, faulty parallelism, and so on. No item contains more than one underlined error.

Now proceed with the following sentences.

C = Correct Sentence

1. There were several letters that he knew he had to write, he told himself that he would write them  

1234

on Wednesday afternoon.
C  
5
2. Whenever I go to a fashion show, I like to get an insight into the latest in women's clothes.  

123

Perhaps even buy something.
C  
5

4
3. I really enjoy sitting in front of a fire blazing away with a special friend on a cold winter's day.  

1234

C  
5
4. The ability to play tennis well demands a strong arm, running and really excellent  

1234

co-ordination.
C  
5
5. Despite the fact that it was a holiday weekend, the line-ups at the ferry terminal  

123

were not all that long.
C  
5

4
6. There is a real problem concerning the amount of garbage we produce which is increasing.  

1234

C  
5
7. Driving in the downtown traffic during the evening rush hour can sometimes be very hard  

123

on one's nerves.
C  
5

4
8. The increased cost of food and clothing were primarily responsible for last month's rise  

123

in the cost of living.
C  
5

4
9. The noise made by the large jet planes flying over our part of the city often  

12

becomes quite upsetting. Especially in the early morning.
C  
5

34
10. I have always been particularly fortunate in my opportunities for travel, when I was only  

123

14 years old, I spent two months travelling in Greece.
C  
5

4

## Section 5: English Usage

Some, though not all, of the ten items found below contain errors in standard English usage, such as the wrong form of a verb, the wrong form of a noun, an incorrectly used preposition, and so on. Circle the number (1, 2, 3, or 4) beneath the underlined portion of the sentence that you think contains an error. If you think the sentence is CORRECT as it appears, circle number 5 under the 'C' at the right of the item. No item contains more than one underlined error.

- |     |  |   |  | C = Correct Sentence |
|-----|--|---|--|----------------------|
| 1.  | In <u>recent</u> years, windsurfing <u>has become</u> one of the <u>most</u> popular <u>summer sport</u> in Canada.                              | C |  |                      |
|     | 1 2 3 4  |   |  | 5                    |
| 2.  | <u>Capital punishments</u> are not justified <u>when</u> the accused <u>has not been</u> proven guilty <u>by means</u> of absolute evidence.     | C |  |                      |
|     | 1 2 3 4  |   |  | 5                    |
| 3.  | The tornado <u>swept over</u> the village <u>with</u> terrible force <u>and</u> left much destruction <u>behind it</u> .                         | C |  |                      |
|     | 1 2 3 4  |   |  | 5                    |
| 4.  | Sonia still <u>has not been</u> able <u>to get use to</u> the fact that <u>in Canada</u> we drive <u>on the right-hand side</u> of the road.     | C |  |                      |
|     | 1 2 3 4  |   |  | 5                    |
| 5.  | <u>I believe that</u> the very best way <u>for an individual to serve the humanity</u> is by <u>becoming a</u> teacher, a scientist or a farmer. | C |  |                      |
|     | 1 2 3 4  |   |  | 5                    |
| 6.  | <u>Incidents</u> that embarrass us <u>when they happen</u> can sometimes be <u>amused over</u> several years <u>later</u> .                      | C |  |                      |
|     | 1 2 3 4  |   |  | 5                    |
| 7.  | <u>Possibly</u> you will think <u>quite different</u> about your job <u>after</u> you come back from <u>your annual vacation</u> .               | C |  |                      |
|     | 1 2 3 4  |   |  | 5                    |
| 8.  | <u>Increased</u> crime in many suburban areas <u>has caused</u> hundreds of <u>families</u> to install <u>burglar alarm systems</u> .            | C |  |                      |
|     | 1 2 3 4  |   |  | 5                    |
| 9.  | <u>Her optimistic attitude</u> is the <u>trait</u> that <u>have always endeared</u> her to her <u>many friends</u> .                             | C |  |                      |
|     | 1 2 3 4  |   |  | 5                    |
| 10. | He <u>complained to</u> the manager <u>on the fact</u> that the salesperson <u>at the ticket counter</u> had been <u>unnecessarily</u> rude.     | C |  |                      |
|     | 1 2 3 4  |   |  | 5                    |

## Appendix G

### Protocol for Administration of the Test of Test-Wiseness

#### DIRECTIONS FOR ADMINISTRATION

Please ensure that you have read the accompanying cover letter to the administering teacher. It contains important general information.

1. Ensure that all students have a SOFT (HB or softer) pencil and eraser.
2. Read the directions regarding how to fill in the bubbles. Ensure that all students understand the procedure.
3. Have the students fill in the background information on the answer sheet:

**Name** - Last name followed by first name with a blank space between

**Sex** - To the right of the name grid

**Birth Date** - Bottom left: Month, Day and Year

**Special Codes** - Bottom middle

**K Column** - Ethnicity

English -	Fill in the "0" bubble
French -	Fill in the "1" bubble
Native -	Fill in the "2" bubble
East Indian -	Fill in the "3" bubble
Chinese -	Fill in the "4" bubble
German -	Fill in the "5" bubble
Italian -	Fill in the "6" bubble
Japanese -	Fill in the "7" bubble
Vietnamese -	Fill in the "8" bubble
Other -	Fill in the "9" bubble

- 2 -

**L column** - "Have you ever had any coaching or specific lessons on how to take a test?"

No, never - Fill in the "0" bubble  
Yes, once or twice - Fill in the "1" bubble  
Yes, three or more times - Fill in the "2" bubble

**M column** "Have you ever practiced writing provincial examinations using previous provincial examinations or questions from these examinations?"

No, never - Fill in the "0" bubble  
Yes, once or twice - Fill in the "1" bubble  
Yes, three or more times - Fill in the "2" bubble

4. Go over the student instructions and the example on page 2 with the students.
5. Begin the test. Please do not allow any discussion among the students. The time provided is a **suggested** time only. Use your discretion to allow additional time if you deem it necessary.
6. Encourage those who seem to be getting frustrated to try and do as well as possible.
7. Following completion of the test, collect all test booklets and answer sheets, both used and unused, and return them to your principal.

THANK YOU!

## Appendix H

### Protocol for Administration of the Language Proficiency Index

To the Administering Teacher:

Enclosed you will find a package of "Student Survey" booklets for your students. The surveys are designed to measure students' "Attitude Toward Tests", "Study Habits", "Achievement Motivation", and ability in selected components of English usage. The first three sections have statements or questions that have no right or wrong answers. Students are to circle the number which indicates how they feel the statement or question applies to themselves. The English usage sections do have right or wrong answers. The students either circle the "5" under the "C" if they think the passage is correct, or circle the number under the error that they have detected.

Please do not discuss or interpret the vocabulary, statements, or questions in the survey with your students. Use your judgement in answering any questions the students may have regarding how to indicate their choice or on any procedural matters, but please do not assist them with interpretation.

#### Directions for Administration

1. Hand out the surveys.
2. Have the students enter their name, school name, and birth date in the boxes at the top of page 1.
3. **IMPORTANT:** Go over the "Instructions for Sections 1, 2, and 3" at the top of page 1. Ensure that the students understand that there are no right or wrong answers in these sections and that they are to answer how they feel. **BEFORE BEGINNING THE SURVEYS**, turn to page 3 and go over the "General Instructions for Sections 4 and 5". Ensure that all students understand that there are right and wrong answers to the questions in these sections, and that they know how to indicate their choice.
4. Begin the survey.
5. As the students are responding, please watch them to ensure that they are not spending too much time on any one item.
6. Please allow as much time as possible for all students to answer each question.

After the survey is completed, please package up all used and unused materials and give them to your principal. He will collect all materials from your school and forward them to us.

Thank you very much for your assistance in this project. If you have any questions about this survey or the overall project, please do not hesitate to contact us at 228-2991 (Rogers) or 228-5298 (Bateson).

## Appendix I

### Preliminary Analysis: The Unit of Analysis using ONEWAY

Table I.15: Results of ONEWAY Analyses Between Schools for the TTW and LPI

Source of Variation	df	Mean Squares	<i>F-Ratio</i>
Absurd Options (TTW)			
Between groups	6	20.87	3.32
Within groups	732	6.28	
Total	738		
Similar Options (TTW)			
Between groups	6	2.99	0.47
Within groups	732	6.43	
Total	738		
Opposite Options (TTW)			
Between groups	6	7.40	1.16
Within groups	732	6.39	
Total	738		
Stem Option (TTW)			
Between groups	6	5.89	0.92
Within groups	732	6.40	
Total	738		
Sentence Structure (LPI)			
Between groups	10	5.45	0.85
Within groups	728	6.41	
Total	738		
English Useage (LPI)			
Between groups	10	14.81	2.36
Within groups	728	6.28	
Total	738		

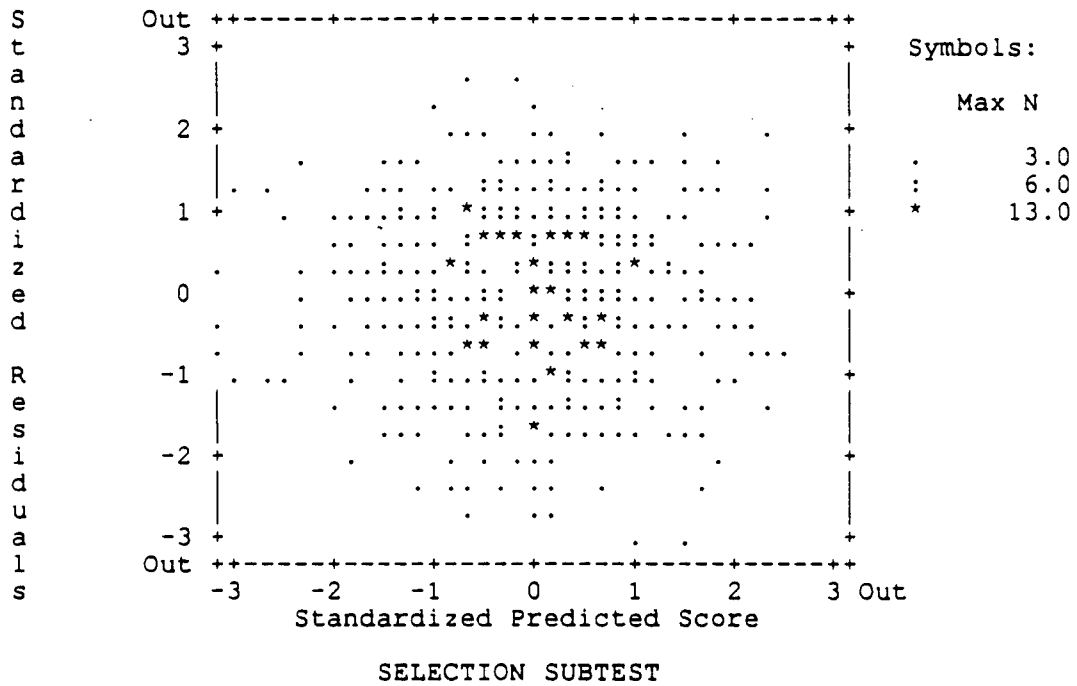
Note:  $0.01 F_{6,732} = 2.82$

$0.01 F_{10,728} = 2.36$

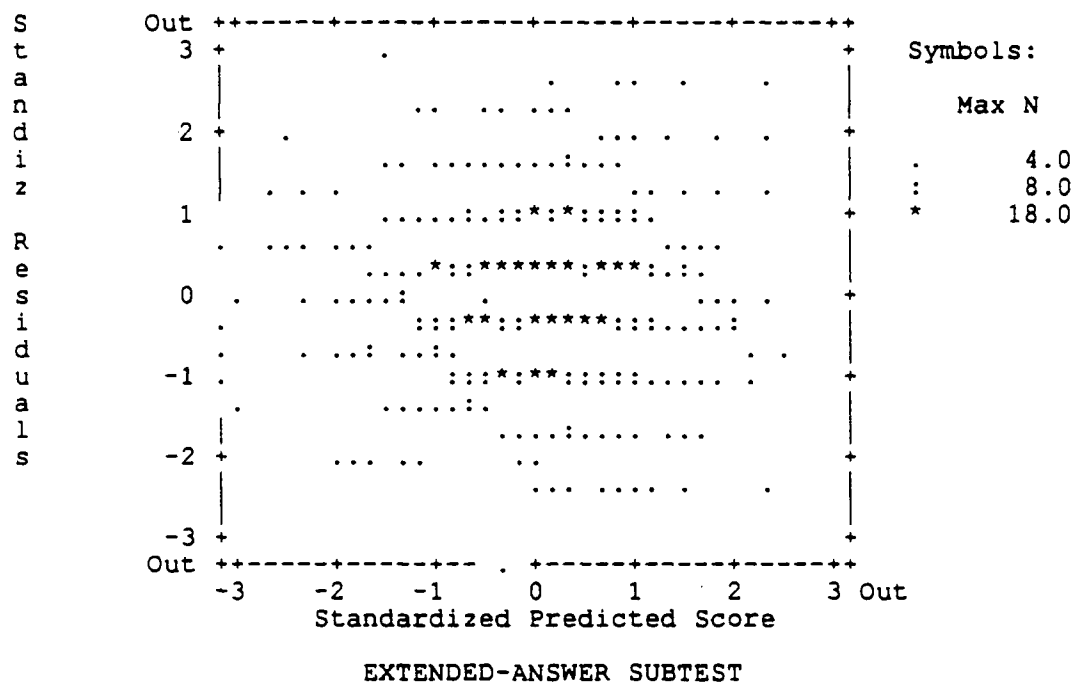
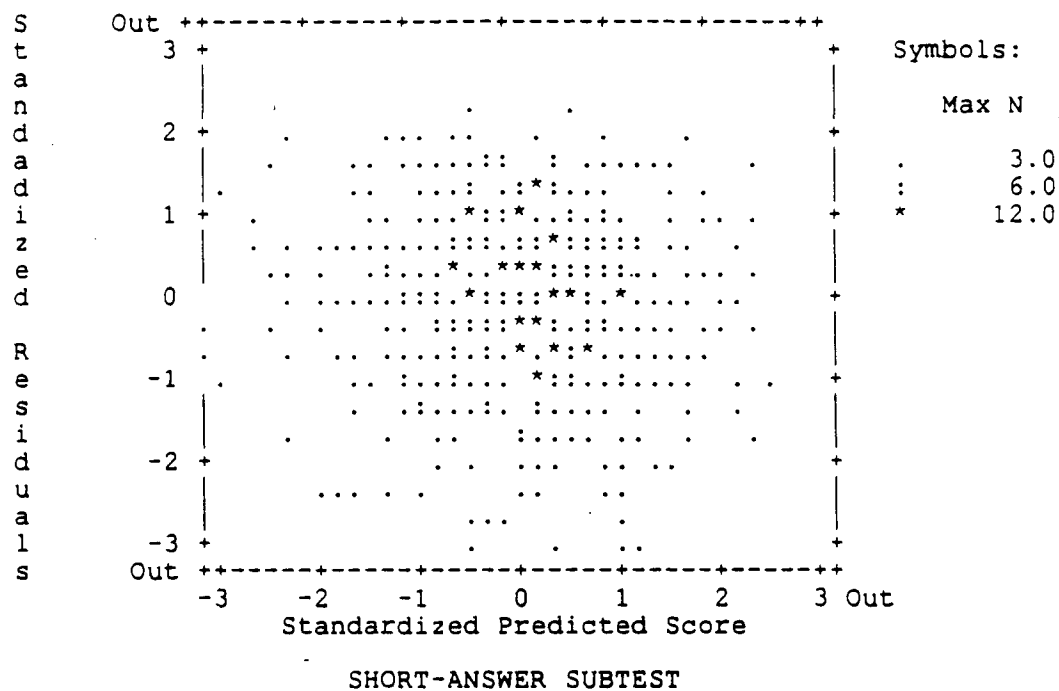
No two groups were found to be significantly different at the 0.05 level using the Tukey method of multiple comparisons.

## Appendix J

### Scatterplots for the Standardized Residuals (n=735)







## Appendix K

### Scatterplots for the Standardized Residuals (n=137)

